DOCUMENT RESUME

ED 198 479 CS 005 850

AUTHOR Carpenter, Patricia A.
TITLE Comprehension Processes in Reading. Final Report.
INSTITUTION Carnegie-Mellon Univ., Pittsburgh, Pa. Dept. of Psychology.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C. Basic Skills Group.
BUREAU NO 6-0388B
PUB DATE 31 Mar 80
GRANT NIE-G-77-0007
NOTE 70p.: Not available in paper copy due to marginal legibility of original document.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS College Students: *Eye Fixations: Higher Education: Memory: Models: *Reading Comprehension: *Reading Processes: *Reading Research: *Recall (Psychology): *Retention (Psychology)

ABSTRACT

Two research studies were conducted to examine short term memory capacity and eye fixations as parts of the reading comprehension process. In the first study, varied aspects of the text and the readers' tasks were examined by monitoring the duration and sequence of readers' eye fixations. The results showed that readers made longer pauses at points of increased processing, such as encoding infrequent words, parsing and representing information in clauses, integrating information from more important clauses, and making inferences. One major outcome of this research was a relatively detailed process model for predicting the gaze duration of college students reading clauses of scientific texts. In the second study, it was hypothesized that individual differences in reading comprehension were caused by differences in working memory capacity. A test with heavy processing and storage demands was devised to measure the trade-off between reading efficiency and the amount of additional information maintained in working memory. When 20 college students read aloud a series of sentences and then recalled the final word of each sentence, their reading span--the number of final words recalled--varied from two to five. This span correlated with three comprehension measures, including verbal SAT and tests involving fact retrieval and pronominal reference. Similar correlations were obtained with a listening span task, showing that the correlation was not specific to reading. (RL)

FINAL REPORT

PROJECT NO. 6-0388B
GRANT NO. NIE-G-77-0007

COMPREHENSION PROCESSES IN READING

Patricia A. Carpenter
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

March 31, 1980

The research reported herein was performed pursuant to a grant with the National Institute of Education, U. S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

U. S. DEPARTMENT OF

HEALTH, EDUCATION AND WELFARE

NATIONAL INSTITUTE OF EDUCATION

BASIC SKILLS

Table of Contents

1. Papers completed. The papers completed under NIE-G-77-0007 are listed below. Copies of papers that are currently in press are included as the second part of this final report. Copies of the other published papers were included in the interim progress reports.

Carpenter, P., and Just, M. Cognitive processes in reading: models based on reader's eye fixations. In A. M. Lesgold and C. A. Perfetti (Eds.), Interactive Processes in Reading. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Daneman, M., and Carpenter, P. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, in press.

Carpenter, P., and Eisenberg, P. Mental rotation and the frame of reference in blind and sighted individuals. Perception and Psychophysics, 1978, 23, 117-124. (copies in 1978 progress report)

Just, M. and Carpenter, P. The computer and eye processing pictures: The implementation of a raster graphics device. Behavior Research Methods and Instrumentation, 1979, 11, 172-176. (copies in 1978 progress report)

Carpenter, P. A., and Just, M. A. Reading comprehension as eyes see it. In M. A. Just and P. A. Carpenter (Eds.), Cognitive Processes in Comprehension. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977. (copies in 1977 progress report)

Just, M. A., and Carpenter, P. A. Inference processes during reading: Reflections from eye fixations. In J. W. Senders, D. F. Fisher, and R. A. Monty (Eds.), Eye Movements and the Higher Psychological Functions. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978. (copies in 1977 progress report).

Carpenter, P. A., and Just, M. A. Eye fixations during mental rotation. In J. W. Senders, D. F. Fisher, and R. A. Monty (Eds.), Eye Movements and the Higher Psychological Functions. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978. (copies in 1977 progress report)

Overview summary. One major innovation of the current project was the use of eye fixations to track reading comprehension processes. It was previously believed that the eye fixations of a reader were not related to the cognitive processes he was executing. The analogy was made to other systems where the rate of input is not related to the underlying processes of interest; for instance, the rate at which coal is shoveled into a furnace may be only roughly related to the rate of the furnace's burning process. This view was challenged by the current project that showed a very close correspondence between the eye fixations and the underlying reading processes. Unlike the furnace example, there tends to be no lag between what the eye fixates and what the mind is processing. Such results are important since they make it possible to use eye fixations to study comprehension.

The goal of the initial experiments was to demonstrate hat readers spent more time on words, phrases and sentences that required more semantic processing. In a typical study, some comprehension process would be made difficult or easy by varying the reading material. The issue was whether the reader's eye fixations showed an effect. In one study, some paragraphs required a more difficult inference and some required a simple inference. An example of a simple inference was:

> The millionaire was killed on a dark and stormy night. The murderer left no clues for the police to trace...

The inference relating the two sentences is simple because the relationship between killed and murderer is direct. Another subject would read the same paragraph, but a slight change in wording made the inference more difficult:

> The millionaire died on a dark and stormy night. The murderer left no clues for the police to trace...

When the reader encounters the word murderer, he must make the inference that the millionaire died an unnatural death, and some person caused his death. In two studies, with a total of 40 college students, it was found that the subjects took longer to read the sentence with murderer in the condition where the inference was difficult. An analysis of their eye fixations indicated that they spent much of this time on the word murderer itself. These data suggest that inference is made relatively soon after first seeing a key word. Equally important for the current research, the eye fixations can be used to identify when this rapid inference process occurs (Carpenter & Just, 1977; Just & Carpenter, 1978).

Other studies showed that _where_ a reader fixates also depends on his comprehension process. This seems to be true of regressive fixations--fixations back to previously read material. For example, one set of studies indicated that skilled readers sometimes look back at the referent of a pronoun, like _he_ or _she_. The regressive fixations that occur during skilled reading often indicate the reader's attempt to fit in what he is currently reading with what he read in previous sentences.

The link between eye fixations and cognitive processes is not specific to reading; another project has indicated that the relation is equally strong in visual problem solving. In these tasks, subjects were required to judge whether two visually presented figures were the same or different. These tasks required the subject to mentally rotate one figure. The decision process was reflected in the sequence and duration of eye fixations (Carpenter & Just, 1978; Just & Carpenter, 1979).

Our general theory is that mental processes cause eye fixations and not the other way around, eye fixations do not initiate the mental processes. To demonstrate this in visual problem solving, we found that blind subjects and blindfolded, sighted subjects show similar patterns of overall response times in these problem solving tasks, even though eye fixations aren't involved (Carpenter & Eisenberg, 1978). An analogous point can be made in reading. The eye fixations in reading reflect comprehension processes that cause the reader to pause longer on a word or to continue on. Good readers tend to have different patterns of eye fixations because their comprehension processes are different. One cannot make a poor reader into a good reader by trying to alter his eye fixations.

More recently, the research goal has been to account for the total time readers spend on particular words, as a function of the text and their reading goals (Carpenter & Just, in press). This research, which is described in detail in the second part, is the first attempt to account for reading times in terms of a detailed model of the comprehension processes. The model that accounts for reading time includes three main processes: encoding a word and retrieving its meaning, parsing, and integrating. The results are extremely promising; the model gives a fairly accurate account of the time skilled readers take to read the various parts of a scientific passage.

The passages are scientific paragraphs excerpted from _Time_ and _Newsweek_. These passages convey information that is sometimes fairly technical, yet the style is readable. In fact, they provide a way for non-specialists to learn about a specialized area.

The passages all follow a somewhat similar format and that can be described by a grammar. For example, the first sentence usually mentions the general topic, a new scientific discovery or process. There is often some background material, such as where, when or how the discovery was made. Definitions are given for new or unusual terms to relate the new term to something the reader already knows. The passage then develops the theme by giving a description of the process. This development consists of a number of subtopics and details. Finally, there is often some statement of the consequence or importance of the finding.

We have related this grammar to the time readers take to process the various parts of the paragraph. The hypothesis is that the time a reader spends on a particular clause will reflect the time it takes him to understand and integrate the new information with what he already has learned from the passage.

In a second aspect of this research, we varied the reader's task by giving him specific reading goals. In the initial experiment, readers were told that they would have to answer some true-false questions about the paragraph. This reading goal is similar to the one used in reading comprehension tests. In addition, we have examined recall and question-answering tasks.

The goal of this research is to account for the time a reader takes to comprehend a passage and to relate it to his later memory for that passage. The theory has three primary components. The first component is word encoding and lexical retrieval. There are two major factors that affect these processes, syllable length and word frequency. Longer units of text take longer to read and, moreover, infrequent words take longer--at least the first time they are encountered. The second component is sentence parsing and representation. This component is influenced by factors such as the number of case-grammatical relations in a sentence. The third component of the theory is how the information in the current sentence is integrated with information already extracted from the passage. Here, the semantic role plays an important role. In particular, details are easy to integrate--because the existing superstructure allows the details to be added on. By contrast, when a new subtopic is introduced, it takes more time to determine its relation to the previous material.

There were an average of 10-15 readers in each task. Their eye fixations were monitored and scored for location and duration. In the data analysis, each passage was divided into a number of sectors corresponding to clauses. The primary dependent measure is the reading time on each sector. The data analysis relates the various factors, syllable length, word frequency, semantic role, and number of propositions, to reading time using multiple regression techniques.

The results show that these factors account for a significant portion of the reading time of scientific passages. Syllable length accounts for one part of reading behavior; in the true-false task, readers spent about 22 msec/letter. Infrequent words (less than 25 per million in Kucera and Frances) add an additional 53 msec/word.

As predicted, different kinds of information (e.g., subtopics vs. details vs. definitions) take different amounts of time to read. For example, in the verification task, our readers spent relatively little time on details. Topics and definitions took considerably longer than details, an extra 85 msec/word. Similarly, subtopics took somewhat longer, 50 msec/word. This translates into a considerable effect. An average topic sentence might be 15 words long--thus, a reader spends an extra 1275 msec on the topic sentence. The extra time spent on a topic might reflect the time it takes a reader to retrieve the correct information from long-term memory to prepare for the subsequent passage. This would be analogous to opening a file in a computer program.

A final interesting aspect of the data is that readers pause at the end of sentences. Sectors corresponding to the end of a sentence take as much as 300 msec longer than sectors that are not at the end. This additional time may reflect the time necessary to integrate the currently read sentence with previous information.

Some of the parameters of the model vary with the task. Others are relatively stable. For example, word frequency, which has its effect on encoding is stable. By contrast, semantic roles are influenced by the readers goals.

Another aspect of the research has examined the cognitive processes that differentiate readers who are very skilled from readers who are less skilled (Daneman & Carpenter, in press). The hypothesis is that individual differences reflect differences in working memory capacity, specifically, the trade-off between processing and storage functions. A poor reader's processes may be inefficient, so that they lessen the amount of additional information that can be maintained in working memory. A test with heavy processing and storage demands was devised to measure this trade-off. Subjects read aloud a series of sentences and then recalled the final word of each sentence. The reading span, the number of final words recalled, varied from 2 to 5 for 20 college students. This span measure correlated with three reading comprehension measures, including verbal SAT and tests involving fact retrieval and pronominal reference. Similar correlations were obtained with a listening span task, showing that the correlation is not specific to reading. The general argument is that beyond the early stages of reading acquisition, the same cognitive processes are responsible for both listening and reading comprehension. Poor readers may also be poor listeners. While this is only an initial investigation, the results are extremely promising.

Implications for education. One goal of basic research is to determine fundamental relations and rules that have wide applicability. For instance, in physics the laws related to force and momentum explain and predict the behavior of physical systems as diverse as automobiles and nuclear reactors. Similarly, one goal of research in basic skills is to describe basic cognitive mechanisms that operate in a variety of domains. It is impossible to specify what all the applications will be, but there appear to be at least two major areas of potential application, communication and remediation.

Research in reading comprehension may suggest how a writer should construct his message for the potential readers. Comprehending an instruction can be a major source of difficulty in performing everyday tasks. We sometimes encounter complex instructions where there is no conceivable purpose for their complexity. This is true in educational settings; for example, complex instructions often occur in tests that are not supposed to be tests of language comprehension. It also occurs in many other aspects of everyday life that aren't formally educational. For example, complex instructions occur in government forms, instruction manuals, and even recipe books. In these situations, the primary purpose of the instruction should be to help people perform correctly and efficiently. The instructions should be written to minimize comprehension difficulties. A theoretical approach to understanding what texts are more difficult for readers will suggest ways of making everyday instructions easier to comprehend.

The second potential area of application is remediation. Basic research in reading should suggest ways of identifying specific reading problems and ways to avoid or eliminate the problems. One interesting suggestion from the current research is that there is a very close relation between reading and general comprehension skills. As Sticht (1972) pointed out, adults who are poor in gaining information from a written text are often equally poor in understanding spoken instructions. If this turns out to be generally true, it could suggest a slightly different focus for reading remediation. Adult reading skills might be improved by focusing on listening skills as well as specific reading skills. In the very least, these results highlight the fact that reading comprehension involves much more than word recognition. More generally, reading research is converging on a set of basic processes that characterize skilled reading. This richer characterization of normal reading may allow a more precise analysis of specific reading problems. Finally, a developmental perspective might have implications for early childhood education. A developmental model of normal comprehension would be useful in determining how particular skills can be fostered at particular points in a child's education.

In sum, basic research in reading might suggest how to optimize communication between the writer and reader and how to identify and improve the skills of poor readers.

Cognitive Processes in Reading:

Models Based on Readers' Eye Fixations

Patricia A. Carpenter

and

Marcel Adam Just

Department of Psychology

Carnegie-Mellon University

Pittsburgh, Pa. 15213

To appear in A. M. Lesgold and C. A. Perfetti (Eds.), Interactive Processes in Reading. Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.

Overview. This paper develops a process model of reading comprehension, as well as a more general theoretical framework to explain why and how reading changes with the task, the text, and the reader. The model focuses on the processes of word encoding, case role assignment and text integration and tries to account for how long readers spend on various parts of a text. Readers make longer pauses at points of increased processing that correspond to encoding infrequent words, integrating information from more important clauses, and making inferences at the ends of sentences. The model predicts the gaze duration of college students on each clause-like unit of scientific text as a function of the involvement of each of these processes. The durations of some of the processes change when the reader's goals change, for example, when the same text is read under different instructions. The flexibility of reading is ascribed to processing goals that control the extent to which the stages of encoding, case role assignment and integration are executed.

Since eye fixations may be an unfamiliar form of data, Table 1 presents an excerpt that may dispel some misconceptions and illustrate some characteristics of eye fixations that motivate the reading model. This table presents a protocol of a college student reading a technical passage about the properties of non-radioactive isotopes. The reader's task was to read a paragraph and then perform a simple true-false comprehension test. The sequence of fixations within each sentence is indicated by the successively numbered fixations above the word being fixated. The duration of each fixation (in msec) is shown below the fixation number.

Insert Table 1 about here

One possible misconception is that readers are selective about which words they will fixate in the previously unread portions of the text. The data here (and most of our other data) indicate that under virtually all non-skimming conditions almost every content word and many function words are fixated. Another possible misconception is that the time spent on all words is approximately equal, once word length is taken into account. The data in Table 1 show this not to be the case. For example, the words radioisotopes and icons were fixated for a long time. This paper will show that the longer fixations are due to longer processing caused by their infrequency and their thematic importance (so that the reader integrates them with previous long-term knowledge). Also, the fixations at the end of each sentence tend to be long. For example, this reader had fixations of 399 msec and 483 msec on medical research at the end of the first sentence, and 450 msec and 650 msec on being produced at the end of the second sentence. These long fixations will be shown to reflect a processing episode that is evoked at the end of a sentence.

The link between eye fixation data and the theory rests on two assumptions. The first, called the immediacy assumption, is that a reader tries to understand each content word of a text as he encounters it, even at the expense of making guesses that sometimes turn out to be wrong. "Understanding" refers to processing at several levels, such as encoding the word, choosing one meaning of it, assigning it to its referent, and determining its status in the sentence and in the discourse. The immediacy assumption posits that the interpretations at all levels of processing are not deferred; they occur as soon as possible, a qualification that will be clarified later.

Table 1

The eye fixations of a college student reading a technical passage in order to later verify true-false statements. The sequence of fixations within each sentence is indicated by the successively numbered fixations above the word being fixated. The duration of each fixation (in msec) is shown immediately below the fixation number.

```
        4     11
       286   466

         1     2     3     5        6        7          8          9
        166   200   167   299      217      268        317        399

     Radioisotopes have long been valuable tools in scientific and medical

                            5
                           183

   10            1    2     3     4              6     8    7      9
  463          317  250   367   416            333   183  450    650

research.  Now, however, four non-radioactive isotopes are being produced.

              4                           8
             366                         183

   1          2    3      5        6          7     9      10     11
  250        200  367    400      216        233   317    283    100

They are called "icons" -- four isotopes of carbon, oxygen, nitrogen and

  12  13
683 150

sulfur.
```

The second assumption, the eye-mind assumption, is that the eye remains fixated on a word as long as the word is being processed. So the time it takes to process a newly-fixated word is directly indicated by the gaze duration. Of course, comprehending that word often involves the use of information from preceding parts of the text, without any backward fixations. So the concepts corresponding to two different words may be compared to each other, for example, while only the more recently encountered word is fixated. The eye-mind assumption can be contrasted with an alternative view that data acquired from several successive eye fixations are internally buffered before being semantically processed (Bouma & deVoogd, 1974). This alternative view was proposed to explain an unusual reading task in which the phrases of a text were successively presented in the same location. However, the situation was unusual in two ways. First, there were no eye movements involved so the normal reading processes may not have been used. Second, and more telling, readers could not perform a simple comprehension test after seeing the text this way. By contrast, several studies of more natural situations support the eye-mind assumption that readers pause on words that require more processing (Just & Carpenter, 1978; Carpenter & Daneman, Note 2). The eye-mind assumption posits that there is no appreciable lag between what is being fixated and what is being processed. In the research to be described, the durations of the individual fixations on a particular word or phrase of the text are cumulated into a single gaze duration. Then the immediacy and eye-mind assumptions are used to interpret the gaze duration data in the development of a model of reading comprehension.

The paper has four major sections. The first briefly describes a theoretical framework for considering the processes and structures in reading. The second section describes the eye fixation experiment in reading. The third section describes the model itself, with subsections describing each component process of the model, and applies the model to a quantitative analysis of the eye fixation results. The fourth section describes mechanisms that allow for changes in reading processes in different situations, and provides results from a second experiment that pinpoint some of the changes.

Theoretical Framework

Reading can be construed as the coordinated execution of a number of processing stages, such as word encoding, lexical access, assigning semantic roles, and relating the information in a given sentence to previous sentences and previous knowledge. Some of the major stages of the proposed model are depicted schematically in Figure 1. The diagram depicts both processes and structures. The stages of reading in the left-hand column are shown in their usual sequence of execution. The long-term memory on the right-hand side is the storehouse of knowledge, including the procedural knowledge used in executing the stages on the left. The working memory in the middle mediates the long-term memory and the comprehension processes. While it is easy to informally agree on the general involvement of these processes in reading, it is more difficult to specify the characteristics of the processes, their interrelations, and their effects on reading performance.

Insert Figure 1 about here

The nature of comprehension processes depends on a larger issue, namely the architecture of the processing system in which they are embedded. While the human architecture is very far from being known, production systems have been suggested as a possible framework because they have several properties that might plausibly be shared by the human system. (Detailed discussions of production systems as models of the human architecture are presented elsewhere, Anderson, 1976; Newell, 1973, 1980.) There are three major properties of particular relevance here.

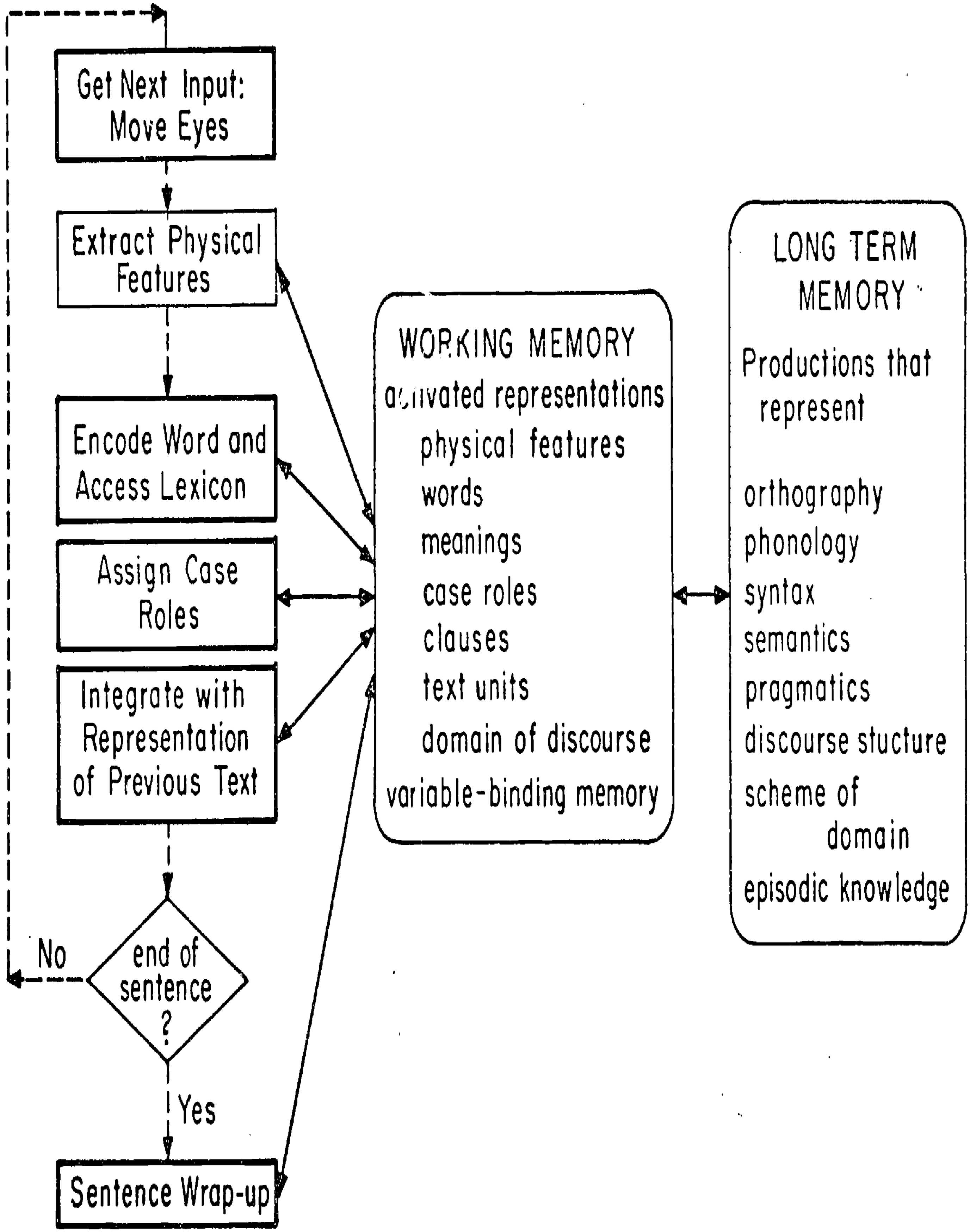


Figure 1. A schematic diagram of the major processes and structures in reading comprehension. Solid lines denote data-flow paths while dashed lines indicate canonical flow of control.

1. Structural and procedural knowledge is stored in the form of condition-action rules, such that a given stimulus condition produces a given action. The productions "fire" one after the other (serially) and it is this serial processing that consumes time in comprehension and other forms of thought. In addition to the serial productions, there are also fast, automatic productions that produce spreading activation among associated concepts (Anderson, 1976; Collins & Loftus, 1975). These automatic productions operate in parallel to the serial productions and in parallel to each other (Newell, 1980). These productions are fast and automatic because they operate only on constants; that is, they directly associate an action with a particular condition (such as activating the concept "dog" on detecting "cat"). By contrast, serial productions are slow because they operate on variables as well as constants; they associate an action with a class of conditions. A serial production can fire only after the particular condition instance is bound to the variable specified in the production. It may be the binding of variables that consumes time and capacity (Newell, 1980). This architectural feature of two kinds of productions permits serial comprehension processes to operate in the foreground while in the background, automatic productions activate relevant semantic and episodic knowledge.

2. Productions operate on the symbols in a limited-capacity working memory. The symbols are the activated concepts that are the inputs and outputs of productions. Items are inserted into working memory as a result of being encoded from the text or being inserted by a production. Retrieval from long-term memory occurs when a production fires and activates a concept, causing it to be inserted into working memory. Long-term memory is a collection of productions that are the repositories of both procedural and declarative knowledge. In the case of reading, this knowledge includes orthography, phonology, syntax and semantics of the language, as well as schemas for particular topics and discourse types (cf. Schank & Abelson, 1977). A new knowledge structure is acquired in long-term memory if a new production is created to encode that structure (Newell, 1980). This occurs if the structure participates in a large number of processing episodes.

One important property of working memory is that its capacity is limited, so that information is sometimes lost. One way that capacity can be exceeded (causing forgetting) is that the level of activation of an item may decay to some subthreshold level through disuse over time (Collins & Loftus, 1975; Hitch, 1978; Reitman, 1974). A second forgetting mechanism allows for processes and structures to displace each other, within some limits (Case, 1978). Heavy processing requirements in a given task may decrease the amount of information that can be maintained, perhaps by generating too many competing structures or by actively inhibiting the maintenance of preceding information. There is recent evidence to suggest that working memory capacity (as opposed to passive memory span) is strongly correlated with individual differences in reading comprehension performance, presumably because readers with greater capacity can integrate more elements of the text at a given time (Daneman & Carpenter, in press).

3. Production systems have a mechanism for adaptive sequencing of processes. The items in the working memory at a given time enable a given production to fire and insert new items, which in turn enable another production, and so on. In this way, the intermediate results of the comprehension process that are placed in working memory can influence or sequence subsequent processing. There is no need for a superordinate controlling program to sequence the mental actions.

The self-sequencing nature of productions is compatible with the model depicted in Figure 1. The composition of each stage is simply a collection of productions that share a common higher-level goal. The productions within a stage have similar enabling conditions and produce actions that serve as conditions for other productions in the same stage. The productions within a stage need not be bound to each other in any other way. Thus the ordering of stages with a production system is accomplished not by direct control-transfer

mechanisms, but an indirect self-sequencing accomplished by one production helping to create the conditions that enable the "next" production to fire.

This architecture permits stages to be executed not only in a canonical order but also in non-canonical orders. There are occasions when some stages of reading appear to be partially or entirely skipped; some stages seem to be executed out of sequence: and some "later" stages sometimes seem to be able to influence "earlier" stages (Levy, this volume). Stages can be executed earlier than normal if their enabling conditions exist earlier than normal. For example, if a context strongly primes a case role, then the case assignment could precede the lexical access of a word. Having read John pounded the nail with a ------, a reader can assign the last word to the instrumental case on the basis of cues provided by the words pound and nail, before encoding hammer. This organization can permit "context effects" in comprehension, where a strong preceding context shortens reading time on a given word or clause. This might occur if a processing stage that is normally intermediate between two others is partially or entirely eliminated. It could be eliminated if the preceding stage plus the context provided sufficient enabling conditions for the later stage. Analogously, a misleading context could lengthen comprehension time by providing elements that enable conflicting processes.

The production system organization can also explain how "later" stages can influence "earlier" stages, so that higher-level schemas can affect word encoding, for example. If the productions of the normally-later stage are enabled earlier than usual, then their outputs can serve as inputs to the normally-earlier stage. The ordering of stages does not have to be entirely reversed to obtain this top-down influence. It may be sufficient for just a portion of the productions of the "later" stage to fire in order to influence the "earlier" stage.

In this view of processing stages, several stages can be executed cotemporaneously in the sense that firings of productions of two or more stages may be interleaved. Consequently, data and control can be transferred back and forth among different stages, somewhat similarly to computer programs organized into co-routines. Co-routines are two or more sub-programs that have equal status (i.e. there is no master-slave relationship); when one co-routine obtains control, it executes until it detects a condition indicating it should relinquish control, and then another co-routine executes, and so on. One interesting difference between co-routines and the production system model is that co-routines generally transfer data between each other only along specified paths, used especially for this purpose. By contrast, productions "transfer" data by placing it in the working memory, so that all processes have access to it. In this sense, the working memory serves as a message center, and communication among stages is by means of the items in working memory. This is distinct from one stage feeding its output directly to another stage.

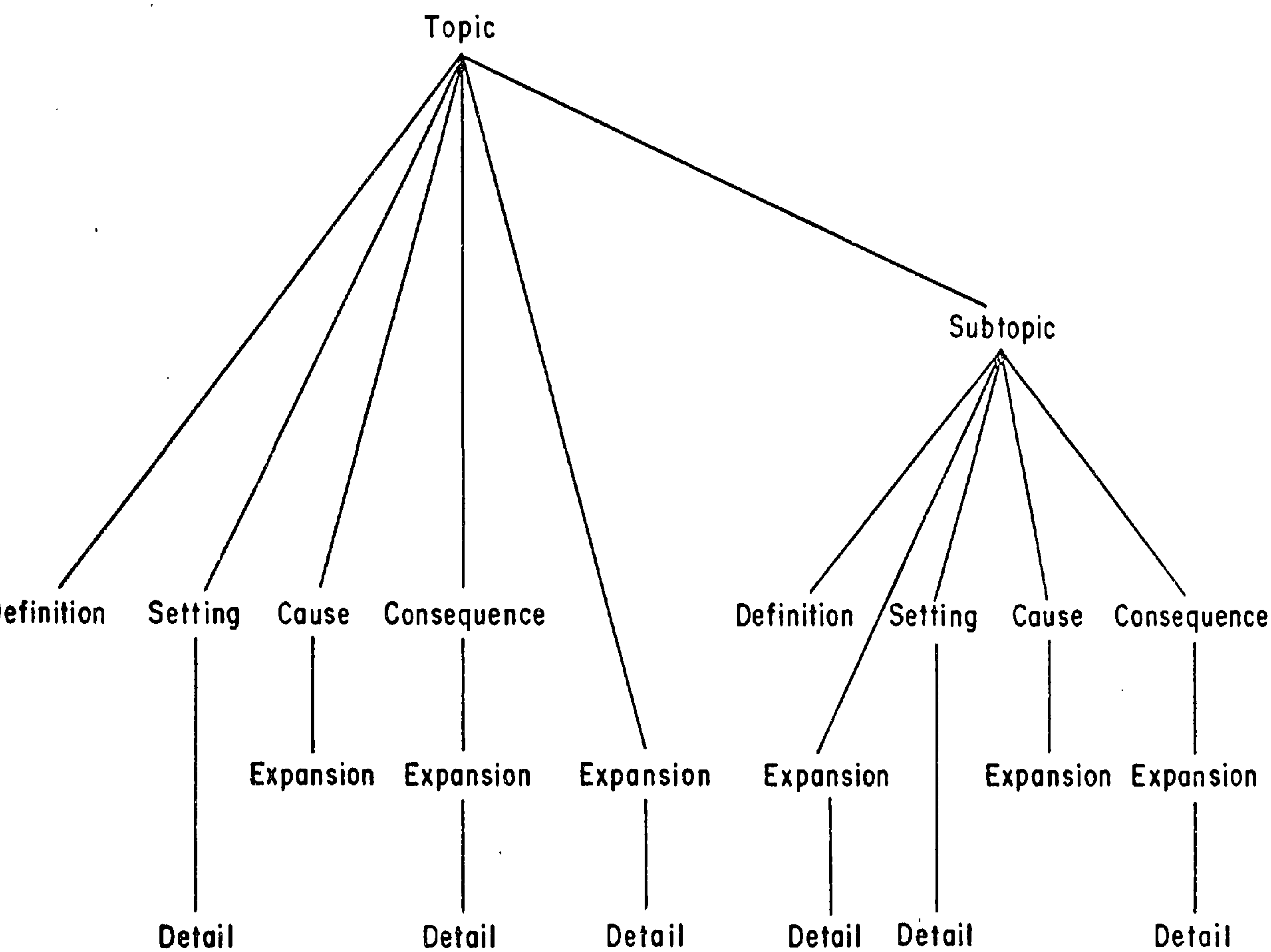
The Research

The texts. This section describes the texts that were used in the reading research because their properties, both local and global, have a large influence on the processing. The global organization of a narrative text has been shown to influence how a reader recalls the text (Kintsch & van Dijk, 1978; Mandler & Johnson, 1977; Meyer, 1975; Rumelhart, 1977; Thorndyke, 1977). The experiment reported below shows that the organization has at least part of its effect when the text is being read. Scientific texts were selected from Newsweek and Time because their content and style is typical of what students read to learn about technical topics. The passages discussed a variety of topics that were generally unfamiliar to the readers in the study. (When readers were asked to rate their familiarity with the topic of each passage, the modal rating was at the "entirely unfamiliar" end of the scale.) There were 15 passages, averaging 140 words each. One of the passages is shown in Table 2. While the texts are moderately well written, they are on

Table 2

A scientific paragraph used in the experiment

Radioisotopes have long been valuable tools in scientific and medical research. Now, however, four non-radioactive isotopes are being produced. They are called "icons" -- four isotopes of carbon, oxygen, nitrogen and sulfur. Each icon has one more neutron in its nucleus than the number usually found in carbon (12), oxygen (16), nitrogen (14), and sulfur (32). The odd neutron gives the isotopes a distinctive magnetic characteristic. These isotopes have some of the same characteristics that made radioisotopes useful for tracing chemical reactions and even powering some of the scientific instruments left on the moon by Apollo astronauts. But the icons do not have the damaging radiation of radioisotopes. Consequently, icons offer good prospects for pharmacological research, for pollution monitoring and for clinical medicine.



igure 2. A schematic diagram of the major text-grammatical categories of information in the scientific aragraphs.

Table 3

A classification of the icon passage

into text grammatical categories

Topic:	Radioisotopes have long been valuable tools in scientific and medical research.
Topic:	Now, however, four non-radioactive isotopes are being produced.
Definition:	They are called "Icons"
Definition:	the four isotopes of carbon, oxygen, nitrogen and sulfur.
Expansion: (of topic)	Each has one more neutron in its nucleus than the number usually found in
Details:	carbon (12), oxygen (16), nitrogen (14), and sulfur (32).
Expansion:	The odd neutron gives the isotopes a distinctive magnetic characteristic.
Subtopic:	These isotopes have some of the same characteristics that made radioisotopes useful for
Detail:	tracing chemical reactions and even powering some of the scientific instruments left on the moon by Apollo astronauts.
Subtopic:	But the icons do not have the damaging radiation of radioisotopes.
Consequence:	Consequently, icons offer good prospects
Details:	for pharmacological research, for pollution monitoring, and for clinical medicine.

the borderline between "fairly difficult" and "difficult" on Flesch's readability scale (1951), with 17 words per sentence and 1.6 syllables per word.

Insert Table 2 and Figure 2 about here

A simplified grammar was developed to categorize the components of the texts. The grammar (shown schematically in Figure 2) classifies the text units, often clauses but sometimes whole sentences or single words, into a structure that is quasi-hierarchical. This abbreviated grammar captures most of the regularities in our short passages (see Vesonder, Note 1, for a more complete grammar for longer scientific passages). The initial sentences generally introduced a topic--a scientific development or event. The beginnings of the passage sometimes gave details of the time, place and people involved with the discovery. Familiar concepts were simply named, while unusual concepts were accompanied by an explicit definition. The main topic itself could be developed through specific examples or through subtopics that were then expanded with further descriptions, explanations and concrete examples. Consequences, usually toward the end of the passage, gave the importance of the event for other applications. Table 3 shows how each text unit or sector in the Icon passage was classified according to these categories (consecutive sectors belonging to the same category have been merged in this table). All 15 passages were segmented into a total of 275 sectors. Each sector was then assigned to one of the five levels by one of the authors. The levels of the grammar were further confirmed by a pre-test involving 16 subjects who rated the importance of each sector in its passage on a 7 point scale. The mean importance ratings differed reliably among the five levels, $F(4,270) = 40.04$, $p<.01$. Specifically, the means decreased monotonically down through the five postulated levels. So the grammar potentially has some psychological reality and its relevance to reading will be demonstrated with the eye fixation data.

Insert Table 3 about here

Method and data analysis. Before the model and data procedures are described in detail, this section presents the data collection and analysis procedures. The readers were Carnegie-Mellon undergraduates who read 2 practice texts followed by the 15 scientific texts in random order. One group of readers was asked to recall each paragraph immediately after reading it, while another group was given about ten sentences to verify as true or false, also immediately after reading the paragraph. The subjects were asked to read naturally, without re-reading or trying to memorize the paragraph, but with a goal of performing well in the recall or verification task. The readers' eye fixations were monitored and recorded as they read the texts.

The paragraphs were presented on a television monitor using upper and lower case and a conventional paragraph layout. The sentences were presented one at a time, and cumulated on the screen to form the paragraph. The subject pressed a button when he came to the end of each sentence, at which point the screen was blanked except for a fixation point indicating where the next sentence began. The reader fixated the point and pressed the button. Then the next sentence was presented along with all of the previous sentences so that the entire paragraph appeared on the screen by the time the subject was reading the last sentence.

The reader's pupil and corneal reflections were monitored relatively unobtrusively by a television camera that was 2.5 feet away. The monitoring system, developed by Applied Science Laboratories, computed the point of regard (rather than eye movement or head movement) every 16.7 msec. If the subject's point of regard was within 1 degree of the fixation point, then the data for that sentence were scorable. This procedure was used to

test whether accuracy of 1 degree was maintained throughout the paragraph. If 1 degree accuracy was not maintained, the data for the entire paragraph were discarded. Due to machine problems, eye fixation data were not obtained from 3 of the subjects in the recall condition and 1 in the verification condition. There were 13 remaining readers in the recall condition and 9 in the verification condition, who on average produced 10 entirely scorable paragraphs, and 5 entire paragraphs for which the data were discarded.

The objective of the data reduction procedure was to convert the 60 observations per second to some dependent measure that could be meaningfully related to reading time. First, there was some reduction as the data were being acquired in real time. A new "fixation" was scored as having occurred if the point of regard changed by more than 1 degree (the size of a three-letter syllable). Furthermore, durations of blinks that were preceded and followed by fixations on the same location were attributed to the reading time on that location. Notice that this treatment of the data ignores traditional fixations, and deals only with the time on a given part of the text, regardless of how many "real" fixations went into that time. The reason for this procedure is that such gaze durations are the most direct behavioral measures to relate to cognitive processes (Just & Carpenter, 1976).

Another program computed the duration of gaze of each subject on each of the 275 sectors. The pooling of all fixation durations on the several words of a sector helps to average over some measurement noise. Since the accuracy of the tracker was about one degree, the measure could be in error by about three letters. By pooling all the fixation durations in a sector, the inaccuracy becomes less important. The mean duration of gaze on each sector (i.e. the average over all the subjects in each experimental condition who produced scorable data for a given paragraph) was computed. The data were fit to the model with a multiple linear regression, in which the independent variables were the factors postulated to affect reading time and the dependent variable was the mean gaze duration on each of the 275 sectors.

The Reading Model

The next subsections describe the five major stages shown in Figure 1: Get next input, Encoding and lexical access. Case role assignment, Interclause integration and Sentence wrap-up. Each subsection describes the processes in that stage together with the factors that should affect the duration of those processes and hence the gaze durations.

Get Next Input. This is the first stage of a cycle that finds information. encodes it and processes it. When the perceptual and semantic stages have done all the requisite processing on a particular word, the eye is directed to land in a new place where it continues to rest until the requisite processing is done, and so forth. The specification of what constitutes "all of the requisite processing" is contained in a list of conditions that must be satisfied before the reader terminates his gaze on the current word and fixates the next one. These conditions include a specification of the goals of normal reading. For instance, one condition may be that a meaning of the word be accessed and another condition may be that a case role be assigned. These conditions can also reflect more specific reading goals. A reader who is trying to memorize a text may have as a condition that the word or phrase be transferred to long-term memory. By setting the conditions appropriately, the reader can adjust his processes to the situation at hand. When the goal conditions for processing a word are satisfied, the resulting action is to Get Next Input.

The command to Get Next Input usually results in a saccade to the next part of the text--one or two words forward. The process that selects the placement of the next forward fixation does not have to be very complex or intelligent. The choice of where to place the next forward fixation appears to depend primarily on the length of the next word

or two to the right of the current fixation (McConkie & Rayner, 1975). The length information, which is encoded parafoveally, is then used to program a rightward saccade. But if only the right margin is visible in the parafovea, then the eye is directed to the first word of the next line, producing a return sweep. In this case the information in peripheral vision is not adequate for accurate targeting. The return sweep is typically too short; the eye often lands on the second word of the new line for a very brief amount of time (50 or 75 msec) and then makes a corrective saccade leftward to the first word of the line (Bayle, 1942). On occasion, a comprehension stage may require a review of previously read text to re-encode it or process it to deeper levels. In those cases, the Get Next Input stage results in a regressive saccade to the relevant portion of the text.

The duration of the Get Next Input stage is short, consisting of the time for a neural signal to be transmitted to the eye muscles. In monkeys, this takes about 30 msec (Robinson, 1972). This duration must not be confused with the typical 150-200 msec latency of a saccade to a visual stimulus that has spatial or temporal uncertainty (Westheimer, 1954). These latencies include stimulus detection, interpretation, and selection of the next fixation target. In normal reading, there is very little uncertainty about direction of the next saccade (it is almost always rightward for forward fixations, except for the return sweeps). Nor is there much uncertainty about distance. On average, the saccade distance may be simply the mean center-to-center distance between words, a distance that does not vary much, relative to the physically possible variation in eye movements. Thus it is reasonable to suppose that the preprogramming time is very short here, consisting usually of a Go signal and the time it takes that signal to be translated into a motor movement, about 30 msec (Robinson, 1972). The actual movements, the saccades, constitute about 5 to 10% of the total reading time. Recent analyses suggest that the saccade itself may destroy the visual persistence of the information from the preceding fixation so that it does not mask the input from the new fixation (Breitmeyer, 1980). Consequently, stimulus encoding may commence soon after the eye arrives at a new location.

<u>Word encoding and lexical access</u>. The reading process involves encoding a word into an internal semantic format. It is assumed that prior to this encoding, the transduction from the printed word to the visual features has already taken place, and that the features have been deposited into the working memory. Perceptual encoding productions use the visual features as conditions; their action is to activate the representation of the word. Once the representation of the word has been sufficiently activated, its corresponding concept is accessed and inserted into working memory. The concept serves as a pointer to a more complete representation of the meaning, which consists of a small semantic network realized as a set of productions. The major nodes of the network are the possible meanings of the word, the semantic and syntactic properties of the meanings, and information about the contexts in which they usually occur (see Rieger, 1979, for a related proposal). The word meanings are represented as abstract predicates, defined by their relations to other predicates.

The productions that encode a word generally trigger on orthographically-based sub-word units, like syllables (Mewhort, & Beal, 1977; Spoehr & Smith, 1973; Taft, 1979). However, there are times when alternative codes, including orthographic, phonological and whole word codes are used (Baron, 1977; Kleiman, 1975; LaBerge & Samuels, 1974). Since the syllable-like encoding unit is believed to be the dominant one, the data should be analyzed in terms of the number of syllables in each word.

The mechanism underlying lexical access is the activation of a word's meaning representation by various sources. There are three ways that a concept's level of activation can be temporarily increased above its base level. One activation mechanism is perceptual encoding; the encoded representation of a word can activate its meaning. A second source is the parallel productions that produce spreading activation through the

semantic and episodic knowledge base of the reader. The third source is activation by the serial productions that do the major computations in all of the stages of processing. When a concept has been activated above some threshold by one or more of these sources, a pointer to its meaning is inserted into working memory. The activation level gradually decays to a subthreshold level, unless some process re-activates it. If the word soon reoccurs in the text while the concept is still activated, lexical access will be facilitated because the activation level will still be close to threshold. When the activation level does decrease, it decreases to an asymptote slightly higher than the old base level. In this way, the system can learn from both local and long-term word repetitions. Frequently used words will have a high base level of activation and consequently will require relatively less additional activation to retrieve them. Thus, frequent words should take less time to access than infrequent words (Morton, 1969). Similarly, the various possible interpretations of each word will have different base activation levels, such that the more common interpretations have higher base activation levels. For example, while the word does has at least two very different meanings, the "third-person singular verb" interpretation would have a higher base activation because it is more common than the "female deer" interpretation (Carpenter & Daneman, Note 2). The more common interpretation would then be accessed faster, since less additional activation would be required to bring the activation level to threshold. This model of lexical access can account for word frequency effects, priming effects, and repetition effects in reading.

Lexical access is complicated by the fact that some words have more than one meaning, so the appropriate interpretation must be selected, or at least, guessed at. When a polysemous word is accessed, the word representation that is retrieved is a pointer to a semantic network that includes the multiple representations. The interpretation that is selected is the one with the highest activation level, and several factors can affect the activation. First, some interpretations start off with a higher activation level; for instance, the "third person singular" interpretation of does has a higher base activation level than the "deer" interpretation. Second, the automatic productions that produce spreading activation can contribute selectively to the activation level of one particular interpretation. The spreading activation can emanate from the preceding semantic and syntactic context, from the reader's knowledge of the domain, and from his knowledge of the discourse style. Third, the output of other stages operating on the same word may activate a particular interpretation. For example, although hammer can be interpreted as a noun or a verb, a sentence context that suggests an instrument to the case role assignment stage (e.g. John hit the nail with a ------.) may help activate the noun interpretation. Fourth, when a word with many highly related meanings occurs in an impoverished context, there may be no single interpretation with higher activation than others, and the superordinate concept may be the selected interpretation of the word. This probably occurs for words that have many closely related interpretations, like get and take.

The selection of only one interpretation of each word, posited by the immediacy assumption, provides a measure of cognitive economy. Selecting just one interpretation allows the system to dampen the activation of the unselected interpretations to keep them from activating their associates. Thus, the contextual effects would remain focused in the appropriate semantic domain. This permits a limited capacity working memory to cope with the information flow in a spreading activation environment that may activate many interpretations and associations for any lexical item. This method of processing also avoids the combinatorial explosion that results from entertaining more than one interpretation for several successive words.

This aspect of the model is consistent with some recent results on lexical access that indicate that although multiple meanings of a word are initially activated, only one meaning remains activated after a few hundred msec. In one experiment, the subjects simultaneously listened to a sentence and pronounced a visually presented word. When an

ambiguous word (rose) was presented auditorally in a syntactic context (e.g., They all rose), the speed of pronouncing a simultaneous visual probe related to either meaning (stood or flower) was faster than in a control condition (Tanenhaus, Leiman & Seidenberg, 1979). In another experiment, the subjects listened to a sentence and performed a lexical decision task on visually presented stimuli. When an ambiguous word (bug) was presented in a semantic context (John saw several spiders, roaches and bugs), the speed of a simultaneous lexical decision related to either meaning (insect or spy) was faster than a control (Swinney, 1979). In both studies, the facilitation of the inappropriate meaning was obtained only within a few hundred msec of the occurrence of the ambiguous word. If the probe was delayed longer, the inappropriate interpretation was no faster than the control. These results suggest that both meanings are available when an ambiguous word is being accessed, but the inappropriate meaning is lost from working memory after a short time.

As the interpretation of the text is constructed, a corresponding representation of the extensive meaning--the things being talked about--is also being built. If a reader cannot determine the referents of the words in a passage, he will find the text difficult to understand. One example of this problem is highlighted in a passage from Bransford and Johnson (1973) that discussed a procedure that involved arranging "things into groups. Of course, one may be sufficient depending on how much there is...." Subjects who weren't given the title "Washing Clothes" thought the story was incomprehensible. The referential representation helps the reader disambiguate referents, infer relations and integrate the text.

The immediacy assumption posits that there is an attempt to relate each content word to its referent as soon as possible. Sometimes this can be done when the word is first fixated, but sometimes more information is required. For example, while the semantic interpretation of a relative adjective like large can be computed immediately, the extensive meaning depends upon the word it modifies (for instance, large insect versus large building). The referent of the entire noun phrase can be computed only after both words are processed. The immediacy assumption doesn't state that the relating is done immediately on each content word, but rather, that it occurs as soon as possible. This is an important distinction that will be made again in the discussion on integrative processes.

Assigning case roles. Comprehension involves determining the relations among words, the relations among clauses, and the relations among whole units of text. This section describes the first of these processes, that of determining the relations among the words in a clause (or in Schank's (1972) terms, determining the dependencies among the concepts). These relations can be categorized into semantic cases, such as agent, recipient, location, time, manner, instrument, action or state (Chafe, 1970; Fillmore, 1968). The case role assignment process usually takes as input a representation of the fixated word, including information about its possible case roles and syntactic properties. For example, hammers tend to be instruments rather than locations or recipients, and information about a word's usual case role can be an important contributor to the assignment process. But this normative information generally is not sufficient to assign its case role in a particular clause. Consequently, the assignment process relies on heuristics that use the word meaning together with information about the prior semantic and syntactic context, as well as language-based inferences. The output of the process is a representation of the word's semantic role with respect to the other constituents in its clause.

Just as certain meanings suggest particular case roles, so too can the context prime a particular case role. Consider the sentence John was interrogated by the ------. The semantic and syntactic cues suggest that the missing word will be an agent, such as detective. The strength of the context becomes evident if the primed case does not occur, for example, John was interrogated by the window. The prior semantic context can precede the affected case assignment by more than a few words. In the sentences The lawyer

wanted to know where in the room John had been interrogated. Mary told him that John was interrogated by the window, the thematic focus of the first sentence on a location alters the interpretation of by and facilitates a locative case role assignment for window.

The specific heuristics that are used in case role assignment have received some attention (see Clark and Clark, 1977, for some examples). Many proposals have the suggestion that readers use the verb as a pivotal source of information to establish the necessary and possible case roles and then fit the noun phrases into those slots (Schank, 1972). But the immediacy assumption posits that the case role assignment for an item preceding the verb is not postponed in anticipation of the verb. Like the lexical access stage, the case assignment stage makes a best guess about a word's case when the word is fixated, rather than making the decision contingent on subsequent words. So the model would not accord any special status to verbs. Another suggested heuristic (that children appear to use) is to assign a sequence consisting of animate noun-verb-noun to the case roles of agent-action-object (Bever, 1970). Like all heuristics, this one sometimes fails, so young children sometimes misinterpret passive sentences (Fraser, Bellugi & Brown, 1963). This heuristic may be employed by adults, but in a modified version that conforms to the immediacy assumption. Rather than waiting for the three major constituents before assigning case roles, the reader should assign an animate noun to the agent role as soon as he encounters it, in the absence of contrary prior context.

The immediate assignment of a case role implies that readers will sometimes make errors and have to revise previous decisions. For example, an adult who assigns the role of agent to an animate noun and then encounters a passive verb will have to revise the agent assignment. (Presumably, young children don't make this revision.) The immediacy of the case assignment process is evident in the reading of sentences like Mary loves Jonathan.... The immediacy assumption suggests that a reader would assign Jonathan the role of recipient; this would turn out to be an incorrect assignment if the sentence continued Mary loves Jonathan apples.

Because case roles are assigned within clauses, the assignment process must include a segmentation procedure to determine clause boundaries within sentences. Sentences can sometimes be segmented into clauses on the basis of explicit markers, such as a subordinating conjunction (e.g., because, when). More often, the reader cannot tell with certainty where one clause ends and another starts until he has read beyond the clause boundary (or potential boundary). A general strategy for dealing with such cases has been suggested, namely to assign a word to the clause being processed, if possible (Frazier & Fodor, 1978). For example, the word soil in the sentence When farmers are plowing the soil... can continue the initial clause (When farmers are plowing the soil it is most fertile) or start a new one (When farmers are plowing the soil is most fertile). The suggested strategy is to continue the initial clause until contrary information is encountered. Interestingly, the strategy discussed by Frazier and Fodor presupposes the immediacy assumption; the segmentation decision arises because case roles are assigned as soon as the words are encountered.

There is no direct mapping between particular case roles and the duration of the assignment process. For example, there is no a priori reason to expect that assignment of instruments takes more or less time than locations. The time for a particular assignment might depend more on the context and properties of the word than on the particular case role being assigned. Detailed specification of the process is not within the scope of this paper; it probably requires a large scale simulation model to examine the complex interactions of different levels of processing.

Interclause integration. Clauses and sentences must be related to each other by the reader in order to capture the coherence in the text. As each new clause or sentence is encountered, it must be integrated with the previous information acquired from the text or with the knowledge retrieved from the reader's long-term memory. Integrating the new sentence with the old information consists of representing the relations between the new and the old structures.

Several search strategies may be used to locate old information that is related to the new information. One strategy is to check if the new information is related to the other information that is already in working memory either because it has been repeatedly referred to or because it is recent (Carpenter & Just, 1977a; Kintsch & van Dijk, 1978). Using this strategy implies that adjacency between clauses and sentences will cause a search for a possible relation. For instance, the adjacent sentences Mary hurt herself. John laughed seem related (John must be a cad) even though there is no explicit mention of the relation. This strategy also entails trying to relate new information to a topic that is active in working memory. This is a good strategy, since information in a passage should be related to the topic.

A second strategy is to search for specific connections based on cues in the new sentence itself. Sentences often contain old information as well as new. Sometimes the old information is explicitly marked (as in cleft constructions and relative clauses), but often it is simply some argument repeated from the prior text. The reader can use this old information to search his long-term text representation and his referential representation for potential points of attachment between the new information and the old (Haviland & Clark, 1974). This second strategy may take more time than the first. In fact, it takes longer to read a sentence that refers to information introduced several sentences earlier, than one that refers to recently introduced information (Carpenter & Just, 1977a).

There are two main points at which integration can occur. First, as each ensuing word of the text is encountered, there is an attempt to relate it to previous information (Just & Carpenter, 1978). Second, a running representation of the clause is maintained, with an updating as each word of the clause is read. This running clause representation consists of the configuration of clause elements arranged according to their case relations. This second type of integration involves an attempt to relate the running clause representation to previous information at each update. Integration occurs whenever a linking relation can be computed. Consider the sentence, Although he spoke softly, yesterday's speaker could hear the little boy's question. The point of this example is not so much that the initial integration of he and speaker is incorrect, but that the integration is attempted at the earliest opportunity. This model implies that integration time may be distributed over fixations on different parts of a clause. Moreover, the duration of the process may depend on the number of concepts in the clause; as these increase, the number of potential points of contact between the new clause and previous information will increase. There is also evidence for integration triggered by the end-of-sentence; this process will be discussed in more detail below.

Integration results in the creation of a new structure. The symbol representing that structure is a pointer to the integrated concepts and this superordinate symbol is then available for further processing. In this way, integration can chunk the incoming text and allows a limited working memory to deal with large segments of prose. The macro-rules proposed by Kintsch and van Dijk (1978) can be construed as productions that integrate.

Integration can also lead to forgetting in working memory. As each new chunk is formed, there is a possibility that it will displace some previous information from working memory. Particularly vulnerable are items that are only marginally activated, usually because they were processed much earlier and haven't recently participated in a production.

For instance, the representation of a clause will decay if it was processed early in a text and was not related to subsequent information. This mechanism can also clear working memory of "lower level" representations that are no longer necessary. For example, the verbatim representation of a previously read sentence may be displaced by the processes that integrated the sentence with other information (Jarvella, 1971). By contrast, the semantic elements that participate in an integration production obtain an increased activation level. This increases the probability that they will become a permanent part of long-term memory.

The main types of interclause relations in the scientific passages correspond to the text-grammatical categories described previously, such as definitions, causes, consequences, examples, and so forth. Text roles that are usually more important to the text and to the reader's goals, like topics or definitions, are integrated differently than less important units, like details. The more central units will initiate more retrievals of relevant previous knowledge of the domain (schematic knowledge), and retrievals of information acquired from the text but no longer resident in the working memory. In addition, more relations will be computed between the semantically central propositions and previous information, because centrality inherently entails relations with many other units. By contrast, details are often less important to the reader's goals and to the text. Moreover, when a detail is to be integrated, the process is simpler because details are often concrete instantiations of an immediately preceding statement (at least in these scientific texts), so they can be quickly appended to information still present in the working memory. Thus, higher level units will take more time to integrate because their integration is usually essential to the reader's goals and because integration of higher units involves more relations to be computed and more retrievals to be made.

The nature of the link relating two stuctures may be explicitly denoted in the text (with connectives like because, therefore, and for instance) or it may have to be inferred on the basis of schematic knowledge of the domain. For example, the causal relation between the sentences Cynthia fell off the rocking horse. She cried bitter tears is inferred from the reader's knowledge about the temporal and causal relation between falling and hurting oneself (Charniak, Note 3).

One cost of immediate interpretation, case role assignment, and integration is that some decisions will prove to be incorrect. There must be mechanisms to detect and recover from such errors. The detection of a misinterpretation often occurs when new information to be integrated is inconsistent with previous information. So misinterpretation detection may be construed as inconsistency detection. For example, the sentence There were tears in her brown dress causes errors initially because the most frequent interpretation of tears is not the appropriate one here and the initial interpretation is incompatible with dress. The eye fixations of subjects reading such garden path sentences clearly indicate that readers do detect inconsistencies, typically at the point at which the inconsistency is first evident (Carpenter & Daneman, Note 2). At that point, they use a number of error-recovery heuristics that enable them to reinterpret the text. They do not start reinterpreting the sentence from its beginning. The heuristics point them to the locus of the probable error. Readers start the backtracking with the word that first reveals the inconsistency, in this case, dress. If that word cannot be reinterpreted, they make regressions to the site of other words that were initially difficult to interpret, like ambiguous words on which a best guess about word meaning had to be made. The ability to return directly to the locus of the misinterpretation and to recover from an error makes the immediacy strategy feasible.

Sentence wrap-up. A special computational episode occurs when a reader reaches the end of a sentence. This episode, called sentence wrap-up, is not a stage of processing defined by its function, but rather by virtue of being executed when the reader reaches the

end of a sentence. The processes that occur during sentence wrap-up involve a search for referents that haven't been assigned, the construction of interclause relations (with the aid of inferences, if necessary), and an attempt to handle any inconsistencies that could not be resolved within the sentence.

The ends of sentences have two important properties that make them especially good places for integration. First, within-sentence ambiguities are usually clarified by the end of the sentence. For example, if a sentence introduces a new object or person whose identity can't be inferred from the preceding context, some cue to their identity is generally given by the end of the sentence. For that reason, if a reader can't immediately determine the referent of a particular word, he can expect to be told the referent or given enough information to infer it, by the end of the sentence. Indeed, readers do use the ends of sentences to process inconsistencies that they can't resolve within the sentence (Carpenter & Daneman, Note 2). The second property is that the end of a sentence unambiguously signals the end of one thought and the beginning of a new one. It can be contrasted with weaker cues that signal within-sentence clause boundaries, such as commas, relative pronouns and conjunctions, that can signal other things besides the end of a clause. Since ends-of-sentences are unambiguous, they have the same role across sentences and they may be processed more uniformly than within-sentence clause boundaries.

There is ample empirical support for the integrative processing at the ends of sentences. Previous eye fixation studies show that when a lexically-based inference must be made to relate a new sentence to some previous portion of the text, there is a strong tendency to pause at the lexical item in question and at the end of sentence that contains it (Just & Carpenter, 1978). Readers were given paragraphs containing pairs of related sentences; the first noun in the second sentence was the agent or instrument of the verb in the first sentence:

1a. It was dark and stormy the night the millionaire was murdered.

1b. The killer left no clues for the police to trace.

In another condition, the integrating inference was less direct:

2a. It was dark and stormy the night the millionaire died.

2b. The killer left no clues for the police to trace.

It took about 500 msec longer to process sentence (2b) than (1b), presumably due to the more difficult inference linking killer to die. There were two main places where the readers paused for those 500 msec, indicating the points at which the inference was being computed. One point was on the word killer, and the other was on the end of the sentence containing killer. Another eye fixation study showed that integration linking a pronoun to its antecedent can occur either when the pronoun is first encountered or at the end of the sentence containing the pronoun (Carpenter & Just, 1977b).

Reading-time studies also have shown that there is extra processing at the end of a sentence. When subjects self-pace the word-by-word or phrase-by-phrase presentation of a text, they tend to pause longer at the word or phrase that terminates a sentence (Aaronson & Scarborough, 1976; Mitchell & Green, 1978). The pause has been attributed to contextual integration processes, similar to the proposed interclause integration process here. Yet another source of evidence for sentence wrap-up processes is that verbatim memory for very recently comprehended text declines after a sentence boundary (Jarvella, 1971; Perfetti & Lesgold, 1977). The model attributes the decline to the interference between sentence wrap-up processes and the maintenance of verbatim information in working memory.

It is possible that wrap-up episodes could occur at the ends of text units smaller or larger than a sentence. For example, the data of Aaronson & Scarborough (1976) suggest that there are sometimes wrap-up processes at the ends of clauses. The decision of when

and if to do wrap-up may be controlled by the desired depth of processing. For example, skimming may require wrap-up only at paragraph terminations, while understanding a legal contract may require wrap-up at clause boundaries. In fact, the clause-boundary effects obtained by Aaronson and Scarborough are sensitive to the subjects' reading goals.

Results

Data Analysis. The purpose of the analysis was to partition reading time into the various processes outlined in the model. Each of the 15 paragraphs was divided into sectors, as in the example in Table 3, producing a total of 275 sectors. The sectors were classified into the categories described by the grammar. A multiple regression analysis computed how the gaze duration on each sector was affected by the variables presumed to influence different stages of the processing. The regression analysis had eight independent variables that coded: the size of a sector, the number of infrequent words it contained, whether it was the last sector in a sentence, and five more variables that coded the text grammatical level of the sector. More precise descriptions of these measures will be presented with the results.

Verification condition results. The analysis indicates how reading time was distributed across the various processing stages when readers anticipated a true-false comprehension test. The eight independent variables were associated with various processing stages, as indicated by the tripartite division of Table 4.

1. Lexical encoding/access and case role assignment. Even though lexical encoding, lexical access, and case role assignment are believed to be separable processes, their durations cannot be measured separately in this experiment. The combined duration of these processes should depend on two factors. The duration should increase with the size of the sector because larger sectors will have more words to encode and more case roles to be assigned. Sector length was measured in number of character spaces since this can easily be done by a computer program. In retrospect, it would have been preferable to measure the length of a sector in syllables since the syllable might plausibly be a unit of lexical encoding (Spoehr, & Smith, 1973). However, the number of syllables is typically highly correlated with the number of letters, with r above .75 in some similar texts we have examined. The second factor affecting the duration of these processes is the number of infrequent words in a sector, because the time taken to access the meaning of infrequent words is longer. "Infrequent" was operationally defined as occurring less than 25 times per million in the Kucera and Francis norms (1967). A word was counted as infrequent only on its first occurrence in the paragraph. These parameters for encoding, access, and case role assignment are shown in the top part of the table. The 22 msec per character indicates the time to encode words, retrieve their meanings and assign case roles. As Table 4 indicates, the first occurrence of an infrequent word added 51 msec to the gaze, suggesting that one or more of encoding, access, and case role assignment takes longer for unusual words. Frequency effects are often attributed to the lexical access process (Glanzer & Ehrenreich, 1979). The additional intercept of 98 msec is a wastebasket parameter. It suggests that there is a minimal duration on a sector corresponding to the minimum encoding, access, and assignment duration.

Insert Table 4 about here

2. Clause Integration. The model posits that integration time should depend on the text-grammatical level of a clause and on the number of concepts it contains. The corresponding independent variable was the interaction of the indicator variables that represented the five text grammatical levels and the variable that represented the number of content words in the sector, with content words defined as in Hockett (1958).

Table 4

Application of the model to the verification condition

	Regression Weight (msec)
<u>Encoding, Lexical Access, &</u>	
<u>Case Role Assignment</u>	
Time per character	22 *
Time per infrequent word	51 *
Intercept	98
<u>Integration</u> (time per content word)	
Topics	96 *
Definitions/Consequences	99 *
Subtopics	57 *
Expansions	49 *
Details	9
<u>Sentence Wrap-up</u>	309 *

*t has $p<.01$

The text-grammatical role of the sector influenced the duration of the integration processes as shown in the middle portion of Table 4. Important sectors, like topics and definitions, had much longer gaze durations than did unimportant sectors, particularly details, when sector length was controlled in an analysis of covariance, $t(269) = 7.67$ and 7.39 respectively, $p<.01$. For example, the additional time spent integrating topics was 96 msec per content word. Consequently, the introductory topic sentence for the icon passage, with 7 content words, required an additional 672 msec for integration. By contrast, the integration time for details was only 9 msec per content word, the only parameter that was not significantly different from zero. The results suggest that details did not receive processing much beyond what was required for encoding and case role assignment. Integration time increases with the number of content words in a sector because more relations are computed. As the number of content words increases, there are more potential relations to other information in the text and to the reader's prior knowledge.

Different text grammatical roles are integrated to different extents. For example, a reader may not relate details to as many different parts of his representation because details are perceived as less important than other kinds of information. In addition, the integration parameter reflects the probability of having to retrieve information from long-term memory. Because details are generally preceded by a relevant superstructure, they can be integrated faster than if they were to be related to information that has to be retrieved from long-term memory.

3. <u>The sentence wrap-up</u> stage should be reflected by an increase in the fixation duration on sectors that were at the ends of sentences. This was coded by another indicator variable. Sentence wrap-up contributes an additional 309 msec to sentence-terminal sectors, as shown in Table 4. This parameter reflects the inference making, consistency checking, and integration that occurs at the end of a sentence. The size of the parameter must be interpreted with caution because the procedure required readers to push a button at the end of each sentence, and the response preparation and execution may have contributed to this parameter. Recently, the end of sentence effect has been obtained in studies that presented the entire paragraph at one time with subjects responding only at the end of the paragraph. The obtained parameter was smaller (71 msec) but still reliable.

The parameters can be used to estimate the reading time for individual sectors. The estimated reading times are compared to the observed reading times for the icon passage in Table 5. The observed and estimated times for this paragraph correlate quite highly $r(21) = .98$, $p<.01$. Over all the paragraphs the correlation is similarly high, $r(273) = .96$, $p<.01$.

Insert Table 5 about here

This analysis can be used to estimate how reading time was distributed over the major processing stages. The proportions will vary with the text, but the icon passage gives an example of the distribution. For the icon passage, the model accounted for about 26.6 sec of the observed reading time of 27.2 sec. About 17.5 sec, or 66% was consumed by the encoding and case role assignment processes. About 3.3 sec, or 12%, was consumed by the integration stage. About 2.5 sec, or 9% was accounted for by the sentence wrap-up processes. Finally, another 2.2 sec, or about 8% was consumed by the base time, which is an intercept parameter covering processes common to all sectors. The time for saccades was not included in the measured gaze durations, but it adds 5 to 10% overhead beyond the 27.2 sec.

Table 5

observed and estimated reading times for the sectors of the icon paragraph

	Observed	Estimated
Radioisotopes have long been valuable tools in scientific and medical research	3073	2857
Now, however, four non-radioactive isotopes are being produced.	2813	2322
They are called "icons"	1116	961
four isotopes of	646	582
carbon	193	260
oxygen	213	260
nitrogen	276	355
sulfur	343	621
Each icon has one more neutron in its nucleus than the number usually found in	2630	2370
carbon (12)	519	379
oxygen (16)	506	379
nitrogen (14)	287	474
sulfur (32)	536	791
The odd neutron gives the isotopes a distinctive magnetic characteristic.	2526	2099
These isotopes have some of the same characteristics that made radioisotopes useful for	2226	2417
tracing chemical reactions	526	767
even powering some of the scientific instruments left on the moon	1666	1606
by Apollo astronauts	873	988
But the icons do not have the damaging radiation of radioisotopes	2750	2448
Consequently, icons offer good prospects	943	1083

Table 5 continued

for pharmacological research	959	802
for pollution monitoring	639	765
and for clinical medicine	910	995

The parameters in Table 4 are all significantly different from zero except for the integration time for details. In addition, the accuracy of the model can be compared to the variability in reading times among sectors. The standard deviation of reading times among the 275 sectors was 917 msec, reflecting the cosiderable variation in their length, text grammatical role and so forth. Compared to this variation, the model fits the data quite well; the standard error of estimate is 265 msec. The model accounts for 92% of the total variance among the 275 means. However, much of the variation among the means is due to differences in the sizes of the sectors. A model that contains only a variable representing the sector size accounts for 84% of the variance, leaving 16% that is unrelated to sector size. Thus the reading model, which accounts for 92% of the total variance, captures 50% of the variance unrelated to sector size.

The model characterizes some of the reading behavior of individual subjects, and not just the average of their distributions of reading time. The median proportion of variance accounted for in the analysis of each subject's data was 77% in the verification task. Moreover, the parameter values for individual subjects stayed within a reasonable range of the parameter value of the subject means. For example, the encoding and case role assignment time was 22 msec for the data averaged over subjects, while the individual subjects' paramters ranged from 19 to 27 msec in the verification task. Infrequent words took longer for all but one subject in the verification task. All readers took longer to integrate the more important sectors in the text grammar. They also spent additional time at the ends of sentences, ranging from 157 to 525 msec. Thus the model applied to the group data is a reasonable index of individual reading performance as well.

Variation in Reading

It is obvious that people read very differently in different situations, but that there is still something that can be identified as "reading" across diverse tasks. A theory of reading should explain the commonalities and differences. The commonality across reading tasks may reflect the fact that all types of reading involve the basic processes of word encoding and lexical access, case role assignment, integration and inference making. What differs between situations is the extent to which each process is evoked. For example, a reader who is checking for logical consistency may devote relatively more processing effort to integration. At the other extreme, some processes may be reduced to a bare minimum. For example, a skimmer may do very little syntactic processing. Since skimmers fixate relatively few words in a sentence, it might be difficult or impossible to establish complete syntactic coherence. These are examples of how the amount of processing devoted to a particular stage can vary across tasks. The model proposes that the amount of processing done of each type of stage is determined by the goals for that situation. The goals determine the conditions that must be satisfied--the kind and amount of information that the system needs before proceeding to the next word or clause. These goals, in turn, reflect the reader's objective in reading, the nature of the text, and the reader's knowledge of the content domain.

Perhaps the most important influence on reading is the reader's objective in reading. This will often determine both what is read, and how it is read. A reader will consult a technical journal for research information, but pick up a novel for entertainment. However, the text's effects can be separated from the reader's objective and both can have a large effect in their own right. The same scientific texts used in the experiment reported above are read differently when the anticipated comprehension test is somewhat different, as the study reported below shows. The change can be localized to a specific process, with the help of the reading model. Eventually, a major part of a theory of reading will be a taxonomy of reading objectives and their effects on each of the processing stages. Presumably the taxonomic approach would start with some of the more common functions of reading, such as reading to learn, to follow instructions, and to be entertained (cf. Sticht,

1977).

The second major influence, the text structure, interacts with the reading process at several levels. Variables such as word frequency and concreteness influence encoding time and readability (Coleman, 1971). Literary devices such as anaphoric reference and intersentential connectives trigger inferences and influence the speed of understanding (Carpenter & Just, 1977b). The current research shows that different parts of a text such as topic, subtopics, and details are read differently. In particular, semantically central units receive more integration, in a way consistent with the model described earlier.

Another major influence on the reading process, aside from the task demands themselves, is the reader's knowledge of the content domain. Many of the comprehension processes depend on the reader's previous knowledge. Encoding and lexical access depend on previous knowledge of vocabulary; case role assignment depends on the semantic knowledge associated with various concepts; integration and inference-making may vary in both extent and direction, depending on the reader's prior knowledge. Reading that is strongly driven by the reader's previous knowledge is called top-down because the higher-level conceptual processes may operate sooner and therefore influence what are normally earlier, bottom-up processes. Thus two people reading the same text with the same objective may read it differently if their previous knowledge of the content area differs.

The Recall Condition

The second experimental condition, in which subjects were asked to recall the passage instead of answering true-false questions about it, provides a test of this view of adjustment of reading goals to meet task demands. A priori, one might guess that readers who will be expecting to recall would attend more to the higher-order structure of the text (like theme statements) in order to help them organize their storage of to-be-recalled material. Also, since the passages contain technical details, readers might spend more time trying to integrate them if they intend to recall the details. By contrast, in the verification condition, the exact details did not have to be integrated completely, since the verification task presented sentences from the text, and so the details were provided at the time of test.

Recall condition results. In the encoding and case role assignment stages, the results for the recall task (Table 5) are very similar to the results for the verification task. Readers spent about the same time for encoding and case role assignment (23 msec per character). Infrequent words had the same effect--adding an average of 51 msec on their first occurrence. The intercept was 93 msec per sector. These results suggest that readers were executing the encoding and case role assignments similarly in the two tasks.

Insert Table 6 about here

The integration stage was sensitive to the differences between the tasks. Readers in the recall condition spent more time on details, 47 msec per word, as opposed to 9 msec in the verification task, $t(20) = 1.93$, $p < .07$. The time spent on details in the recall condition was significantly greater than zero. Finally, the sentence wrap-up parameter (192 msec) was still fairly substantial.

The model's fit was as good as for the verification task. All of the models' parameters were significantly different from zero. The standard error of estimate was 2[illegible]3 msec, which is reasonable compared to the standard deviation of 947 msec among the reading times for the sectors. The regression accounted for 92% of the results for the

Table 6

Application of the model to the recall condition

	Regression Weight (msec)
Encoding, Lexical Acces, & Case Role Assignment	
Time per character	23 *
Time per infrequent word	51 *
Intercept	93
Integration (time per content word)	
Topics	65 *
Definitions/Consequences	106 *
Subtopics	81 *
Expansions	76 *
Details	47 *
Sentence Wrap-up	192 *

*t has p<.01

recall means. In general, the fit was comparable in the two conditions.

In summary, the two tasks show striking similarities in the encoding and case role assignment stages. Both tasks resulted in careful reading of about the same rate, 315 words per minute (wpm) for the verification task and 295 wpm for the recall task. In both tasks, integration time decreased as the role varied from an important one (like a topic or definition) to a less important one (like a detail). However, there were some differences in the integration stage. Readers in the free recall condition spent more time integrating details with the superordinate structure than did readers in the verification condition. The recall subjects were probably spending more time on integration because they knew they would later have to produce some of them, rather than merely recognize them. Thus, their criterion for integration was adjusted to meet their objectives.

Skimming. In an attempt to strongly manipulate reader's task, five additional readers were asked to skim each paragraph and recall it immediately afterward. Four of the readers were unable to skim under these conditions, as indicated by their slow reading rate. However, there was one reader who did skim at 925 wpm. A protocol, shown in Table 7, shows a very different behavior from the careful reading performance in the other conditions. Her fixations are shorter and they are selective. She did not fixate every word or even every clause; in particular, she skipped details--like the numbers in line 5 and the list of applications for radioisotopes in line 8. She spent more time on important sectors, such as the first line and line 10, which mentions that icons lack radiation.

Insert Table 7 about here

Her recall also indicates that she did not process details; it contains only the more important text grammatical units. A typical example is her recall of the icon passage:

> This paragraph was talking about a kind of substitute for radioactive isotopes called icons and how they have something in them that gives them a magnetic, kind of magnetic charge and how they're being used in different things. And they're not as dangerous as a radioactive isotope is.

While the skimming data are only preliminary, they suggest that skimming also may rely on the ability to discriminate between various text grammatical roles and to selectively process the units that are relevant to the reader's objective.

Memory. Memory performance should reflect, in part, how well each part of the text was integrated during comprehension. Clauses that are integrated to a greater degree are more likely to become part of long-term memory and should be recalled better. There are two factors that determine the extent to which a unit of text is integrated. First, those parts of the text structure that receive more integration time, like topics and definitions, should be recalled better. Second, if a clause is referred to several times in the passage, either explicitly or implicitly, then it may be involved in an integration episode after each mention. Hence, recall of a given unit should be partially explained by integration time and by the number of times that a unit is referred to in the text. This account resembles the one proposed by Kintsch and van Dijk (1978), except that the current account uses integration parameters estimated from the reading fixations.

The proportion of the 16 subjects that recalled each of the 275 sectors was analyzed as a function of two predictor variables: the integration parameter for that type of text unit (derived from the gaze-duration analysis) and a count of the number of times a sector was referred to in the passage after its initial occurrence. Both variables produced reliable

Table 7

The eye fixations of a college student skimming the passage. The
equence of fixations within each sentence is indicated by the successively
umbered fixations above the word being fixated. The duration of each
ixation (in msec) is shown immediately below the fixation number.

```
     1          2              3           4              5    6        7
   (183)      (133)          (217)       (100)          (267)(67)     (116)
   Radioisotopes have long been valuable tools in scientific and medical

 8                 1                   2              3      4
(117)             (83)               (167)          (184)  (150)
esearch.  Now, however, four non-radioactive isotopes are being produced.

      1          2                    3         4           5
    (183)      (183)                (200)     (133)       (217)
ey are called "icons" -- four isotopes of carbon, oxygen, nitrogen, and

  6           1          2          3          4              5
 (67)       (183)      (67)       (83)       (150)          (200)
ulfur.  Each icon has one more neutron in its nucleus than the number

                              8         7              6
                            (133)     (150)          (216)
sually found in carbon (12), oxygen (16), nitrogen (14), and sulfur (32).

      1         2                       3        4
    (200)     (116)                   (217)    (183)
e odd neutron gives the isotopes a distinctive magnetic characteristic.

      1         2              3               4            5
    (201)     (67)           (67)           (150)        (200)
ese isotopes have some of the same characteristics that made radioisotopes

                                        6        7      8
                                      (217)    (216)  (217)
seful for tracing chemical reactions and even powering some of the

    11   10          12   9         13                  1    2
  (133)(183)        (83)(183)      (200)              (200)(117)
ientific instruments left on the moon by Apollo astronauts.  But the icons

                      8
                    (233)

     4       3       5      6           7         1       2    3
   (167)   (133)   (217)  (67)        (200)     (183)   (284)(133)
 not have the damaging radiation of radioisotopes.  Consequently, icons

       4      5                6     7        9        8
     (233)  (67)             (200) (133)    (200)    (216)
fer good prospects for pharmacological research, for pollution monitoring

              10
            (284)
d for clinical medicine.
```

effects, $t(272) = 5.60$, $p<.01$, for the integration time parameter and $t(272) = 5.79$, $p<.01$, for the number of repetitions. Thus certain aspects of comprehension are at least partial determinants of recall.

Text units that were higher in the grammar tended to be recalled better (see Figure 3), replicating the text-role effects found with other types of texts (Meyer, 1975; Thorndyke, 1977). The model partially explains this result in terms of the processes that occur during comprehension. In addition, retrieval processes also may play a role in this effect. For example, there may be many paths from less important concepts that lead to topics, but not vice versa. There may also be response output effects. Furthermore, a complete model of recall will have to consider how the particular facts and concepts in the text map onto the reader's previous knowledge. While the passages used were generally unfamiliar, particular facts surely differed in their familiarity and this could affect recall (Spilich, Vesonder, Chiesi & Voss, 1979).

Insert Figure 3 about here

The verification results were not amenable to this kind of analysis. Each passage had been followed by 4 to 6 true and 4 to 6 false probes, with each true probe containing information from several text units. The true probes were verified fairly accurately, with a mean accuracy of 75% correct and an average response time of 4040 msec.

In summary, the model shows that an account of the comprehension process can be used to partially account for recall performance. To totally explain recall will require a precise account of the role of prior long-term knowledge and the role of retrieval and reconstruction processes in recall.

Discussion

The results support several aspects of the reading model and the general theoretical framework. At the encoding and case assignment stages, clauses with infrequent words took more time to process. At the integration stage, units that were higher in the text grammar received longer gaze durations. This greater duration reflects a more extensive integration process, both in terms of the number of times information would be retrieved from long-term memory and in the number of relations that would be computed. Details, because they are relatively unimportant to the discourse structure, took less time to be integrated than did other categories of information. Verification readers may have had a lower criterion for integration of details since they only had to recognize them. However, readers in the recall condition paid more attention to details. Finally the end of a sentence can trigger additional processing, including inference making and integration. The differences in criteria for the two tasks were apparent only at the integration stage; however, the theory suggests that encoding, case role assignment, integration and sentence wrap-up may all vary if the reader's goals change sufficiently. The model and the analytic techniques make it possibly to identify where tasks have their effects. Thus, it is possible to go beyond global statements of task effects and to quantify where and how reading objectives influence reading.

Eye fixations, reading, and listening. The reported research demonstrates that readers do not distribute their gazes uniformly during reading; they look longer at text units where they are doing more processing. This selective allocation of processing resources during reading can be compared to analogous mechanisms in listening. Although a listener cannot control the rate of input or re-listen to the spoken discourse, there are mechanisms that functionally resemble the pacing and rereading role of eye fixations. For example, a listener's echoic memory allows him to store some part of the auditory message

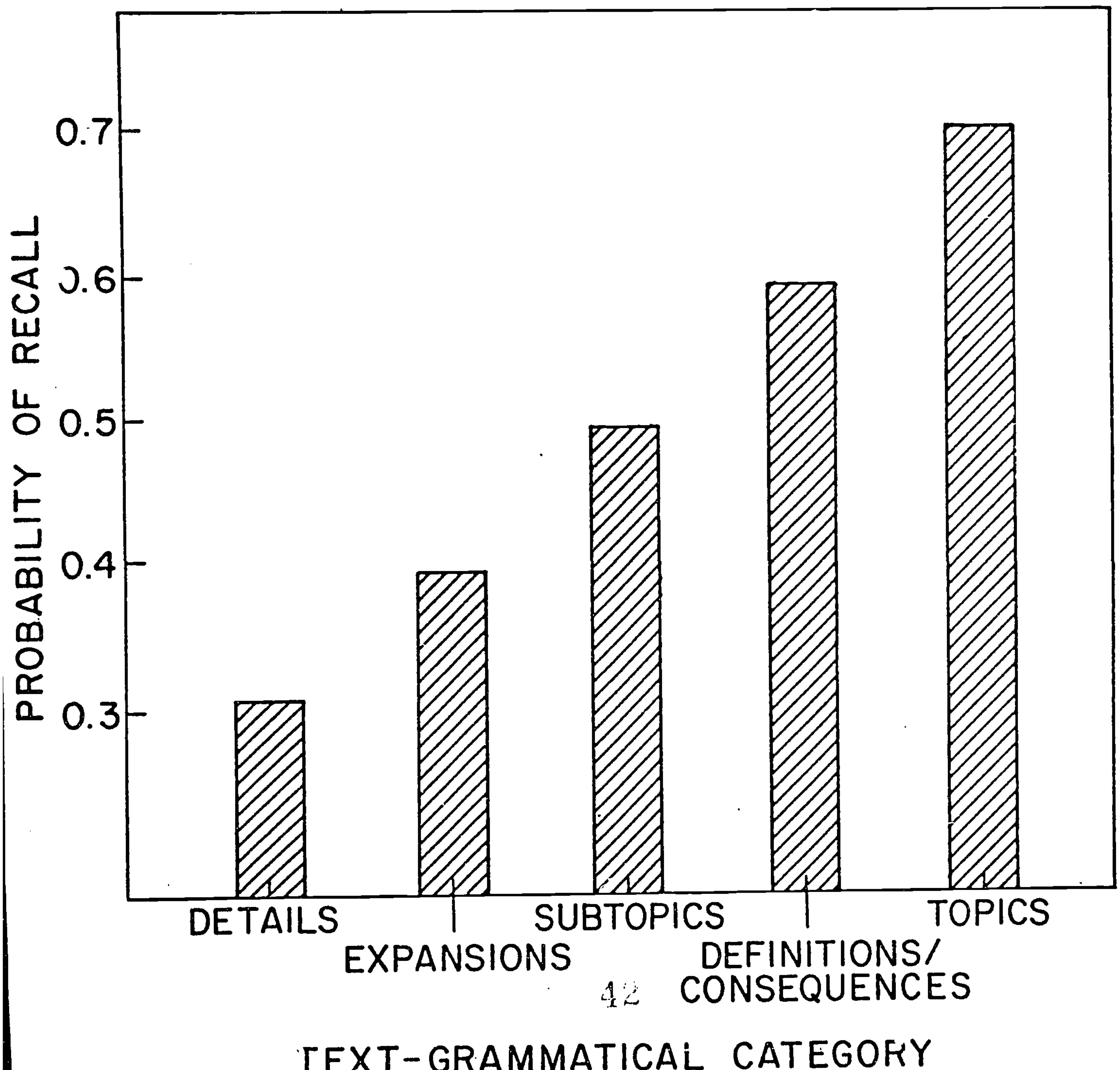


Figure 3. The probability of recall of each text-grammatical category.

temporarily, at least partially compensating for the large external memory available to the reader. The listener may also use the speaker's pauses to do extra processing at points of high computational demand. Moreover, the listener is aided in identifying important text roles and relations by the speaker's repetition, stress and gesture. To the extent that these devices allow a listener some selectivity in processing, auditory comprehension may resemble its reading counterpart.

Although reading and listening comprehension have some analogous mechanisms, they may not produce equivalent comprehension. The theoretical framework accounts for this with its notion of variable goals for each stage. The goals for the comprehension stages for reading may differ from those in listening because the processes may not be able to perform to the same specifications in the two modalities. For example, a listener might not be able integrate to the same extent as a careful reader. If the auditory material were presented too quickly, the listener might allow inconsistencies to go unchecked or else risk missing the incoming information. With sufficiently sensitive methodologies, it should be possible to find systematic differences between the results of reading and listening to the same text.

The immediacy assumption. The model's ability to account for gaze durations in terms of the comprehension processes provides some validation for the immediacy and eye-mind assumptions. Readers interpret a word while they are fixating it, and they continue to fixate it until they have processed it as far as they can. This kind of processing eliminates the memory load and computational explosion that would result if a reader kept track of several possible meanings, case roles, and referents for each word and computed the final interpretation at the end of a clause or sentence. This architectural feature also allows a limited capacity processor to operate on a large semantic network without being bombarded by irrelevant associations. After a single interpretation has been selected, the activation of the unselected meanings can be dampened to their base levels so that they will not activate their semantic associates any further. This minimizes the chances that the reader will be conceptually driven in many directions at the same time.

The cost of this kind of processing is fairly low because the early decisions usually are correct. This is accomplished by taking a large amount of information into account in reaching a decision. The processes have specific heuristics to combine semantic, syntactic and discourse information. Equally important, the processes operate on a data base that is strongly biased in favor of the common uses of words and phrases, but one that also reflects the effects of local context. The cost is also low because the reader can recover from errors. It would be devastating if there were no way to modify an incorrect interpretation at some later point. However, there are error-recovery heuristics that seem fairly efficient, although the precise mechanisms are only now being explored (Carpenter & Daneman, Note 2).

The fact that a reader's heuristics for interpreting the text are good explains why the "garden path" phenomenon isn't the predominant experience in comprehension; it only happens occasionally. Perhaps the most common, everyday "garden path" experiences occur when reading newspaper headlines, for example, Carter views discussed and Judge admits two reporters. The incorrect initial interpretations occur because headlines are stripped of the syntactic and contextual cues that guide the processing of normal text. Similarly, many jokes and puns explicitly rely on the contrast between two interpretations of an ambiguous word or phrase (Shultz & Horibe, 1974). Even garden path sentences sometimes seem funny. The humor in all of these cases resides in the incongruity between the initial interpretation and the ultimate one. Garden path sentences are also infrequent because writers usually try to avoid ambiguities that might encourage or allow incorrect interpretations. These kinds of sentences are useful tools for studying comprehension because they indicate where the usual comprehension strategies fail. But the fact that

they are not frequent indicates that a reader's heuristics usually are sufficient.

The current research demonstrates that specific fixation patterns can be related to specific comprehension processes (cf. also Carpenter & Just, 1977a; Just & Carpenter, 1976, 1978). This approach has its roots in reading research earlier in the century when many researchers considered eye fixations to be excellent indicators of comprehension (Buswell, 1920, 1937; Dearborn, 1906; Huey, 1908; Judd & Buswell, 1922). Eye fixations were correctly interpreted as the result, not the cause of reading processes. A reader who had difficulty comprehending a passage would show an unusual pattern of eye fixations because of his cognitive problems. The primary measures used in those studies were mean fixation duration and mean number of fixations, averaged over an entire passage. It was found that both measures increased for texts that were intuitively classified as difficult. But there was little systematic effort to relate pauses to particular parts of a text. This relation evident in this earlier research and the current results, shows that eye fixations are a useful window through which to view comprehension processes. They offer a means of attaining a goal that Huey (1908, p. 6) set for psychology:

> ...to completely analyze what we do when we read would be almost the acme of a psychologist's achievements, for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilization has learned in all its history.

Reference Notes

1. Vesonder, G. T. The role of knowledge in the processing of experimental reports. Unpublished doctoral dissertation, University of Pittsburgh, 1979.
2. Carpenter, P. A., & Daneman, M. Lexical access and error recovery in reading: A model based on eye fixations. Carnegie-Mellon University, Unpublished manuscript, 1980.
3. Charniak, E. Toward a model of children's story comprehension Cambridge, Mass.: MIT Artificial Intelligence Laboratory. Technical Report No. 266, 1972.

References

Aaronson, D., & Scarborough, H. S. Performance theories for sentence coding: Some quantitative evidence. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 56-70.

Anderson, J. R. Language, memory, and thought. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Baron, J. Mechanisms for pronouncing printed words: Use and acquisition. In D. LaBerge and S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Bayle, E. The nature and causes of regressive movements in reading. Journal of Experimental Education, 1942, 11, 16-36.

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), Cognition and the development of language. New York: John Wiley & Sons, 1970.

Bouma, H., & deVoogd, A. H. On the control of eye saccades in reading. Vision Research, 1974, 14, 273-284.

Bransford, J. D., & Johnson, M. K. Considerations of some problems of comprehension. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

Breitmeyer, B. G. Unmasking visual masking: A look at the "why" behind the veil of "how". Psychological Review, 1980, 87, 52-69.

Buswell, G. T. How adults read. Supplementary Educational Monographs, 45, Chicago, Illinois: University of Chicago, 1937.

Carpenter, P. A., & Just, M. A. Integrative processes in comprehension. In D. LaBerge and S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977a.

Carpenter, P. A., & Just, M. A. Reading comprehension as eyes see it. In M. A. Just and P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977b.

Case, R. Intellectual development from birth to adulthood: A neo-Piagetian interpretation. In R. Siegler (Ed.), Children's thinking: What develops? Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Chafe, W. L. Meaning and the structure of language. Chicago: University of Chicago Press, 1970.

Clark, H. H., & Clark, E. V. Psychology and language. New York: Harcourt, Brace and Jovanovich, Inc., 1977.

Coleman, E. B. Developing a technology of written instruction: Some determiners of the complexity of prose. In E. Z. Rothkopf and P. Johnson (Eds.) Verbal learning research and the technology of written instruction. New York: Teachers College Press, Columbia, University, 1971.

Collins, A. M., & Loftus, E. F. A spreading activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Daneman, M., & Carpenter, P. A. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, in press.

Dearborn, W. The psychology of reading. Columbia University Contributions to Philosophy and Psychology. New York: The Science Press, 1906.

Fillmore, C. J. The case for case. In E. Bach and R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart & Winston, Inc., 1968.

Flesch, R. F. How to test readability. New York: Harper, 1951.

Fraser, C., Bellugi, U., & Brown, R. Control of grammar in imitation, comprehension, and production. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 121-135.

Frazier, L., & Fodor, J. The sausage machine: A new two-stage parsing model. Cognition, 1978, 6, 291-325.

Glanzer, M. & Ehrenreich, S. L. Structure and search for the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 381-398.

Haviland, S. E., & Clark, H. H. What's new? Acquiring new information as a process in comprehension. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 512-521.

Hitch, G. J. The role of short-term working memory in mental arithmetic. Cognitive Psychology, 1978, 10, 302-323.

Hockett, C. F. A course in modern linguistics. New York: Macmillan Company, 1958.

Huey, E. B. The psychology and pedagogy of reading. New York: Macmillan Company, 1908, reprinted Cambridge, Mass: The MIT Press, 1968.

Jarvella, R. J. Syntactic processing of connected speech. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 409-416.

Judd, C. H., & Buswell, G. T. Silent reading: A study of the various types. Supplementary Educational Monographs, 23, Chicago: University of Chicago, 1922.

Just, M. A., & Carpenter, P. A. Eye fixations and cognitive processes. Cognitive Psychology, 1976, 8, 441-480.

Just, M. A., & Carpenter, P. A. Inference processes during reading: Reflections from eye fixations. In J. W. Senders, D. F. Fisher, and R. A. Monty (Eds.), Eye movements and the higher psychological functions. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Kleiman, G. M. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 323-339.

Kucera, H., & Francis, W. N. Computational analysis of present-day American English. Providence, Rhode Island: Brown University Press, 1967.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Levy, B. A. Interactive processes during reading. In A. M. Lesgold and C. A. Perfetti (Eds.), Interactive processes in reading. hillsdale, N.J.: Lawrence Erlbaum Associates, this volume.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. Cognitive Psychology, 1977, 9, 111-151.

McConkie, G. W., & Rayner, K. The span of the effective stimulus during a fixation in reading. Perception & Psychophysics, 1975, 17, 578-586.

Mewhort, D., & Beal, A. L. Mechanisms of word identification. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 629-640.

Meyer, B. The organization of prose and its effect on recall. Amsterdam: North-Holland, 1975.

Mitchell, D. C., & Green, D. W. The effects of context and content on immediate processing in reading. Quarterly Journal of Experimental Psychology, 1978, 30, 609-636.

Morton, J. Interaction of information in word recognition. Psychological Revie , 1969, 76, 165-178.

Newell, A. Production systems: Models of control structures. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

Newell, A. Harpy, production systems and human cognition. In R. Cole (Ed.), Perception and production of fluent speech. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Perfetti, C. A., & Lesgold, A. M. Discourse comprehension and sources of individual differences. In M. A. Just and P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Reitman, J. S. Without surreptitious rehearsal, information in short-term memory decays. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 365-377.

Rieger, C. J. Five aspects of a full-scale story comprehension model. In N. V. Findler (Ed.), Associative networks. New York: Academic Press, 1979.

Robinson, D. A. Eye movements evoked by collicular stimulation in the alert monkey. Vision Research, 1972, 12, 1795-1808.

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge and S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Scarborough, D. L., Cortese, C. & Scarborough, H. S. Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 1-17.

Schank, R. C., & Abelson, R. P. Scripts, plans, goals and understanding: An inquiry into human knowledge structures. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Schank, R. C. Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 1972, 3, 552-631.

Shultz, T., & Horibe, F. Development of the appreciation of verbal jokes. Developmental Psychology, 1974, 10, 13-20.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. Text processing of domain-related information for individuals with high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 275-290.

Spoehr, K. T., & Smith, E. The role of syllables in perceptual processing. Cognitive Psychology, 1973, 5, 71-89.

Sticht, T. G. Comprehending reading at work. In M. A. Just, and P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Swinney, D. A. Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 645-659.

Taft, M. Recognition of affixed words and the word frequency effect. Memory & Cognition, 1979, 7, 263-272.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 427-440.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Westheimer, G. H. Eye movement responses to a horizontally moving visual stimulus. Archives of Opthamology, 1954, 52, 932-943.

Footnote

The order of authors was decided by the toss of a coin. The research was supported in part by grant NIE-G-77-0007 from the National Institute of Education and grant MH-29617 from NIMH. We thank J. Anderson, L. Noordman, V. Noordman-Vonk, L. Reder and H. Simon for their comments on earlier drafts of the manuscript.

Individual differences in working

memory and reading

Meredyth Daneman

and

Patricia A. Carpenter

Department of Psychology

Carnegie-Mellon University

Pittsburgh, Pa. 15213

(412) 578-2792

Journal of Verbal Learning and Verbal Behavior, in press

Abstract

Individual differences in reading comprehension may reflect differences in working memory capacity, specifically in the trade-off between its processing and storage functions. A poor reader's processes may be inefficient, so that they lessen the amount of additional information that can be maintained in working memory. A test with heavy processing and storage demands was devised to measure this trade-off. Subjects read aloud a series of sentences and then recalled the final word of each sentence. The reading span, the number of final words recalled, varied from 2 to 5 for 20 college students. This span correlated with three reading comprehension measures, including verbal SAT and tests involving fact retrieval and pronominal reference. Similar correlations were obtained with a listening span task, showing that the correlation is not specific to reading. These results were contrasted with traditional digit span and word span measures which do not correlate with comprehension.

Many theorists have suggested that working memory capacity plays a crucial role in reading comprehension (cf. Just & Carpenter, in press; Kintsch & van Dijk, 1978); however, traditional measures of short-term memory, like digit span and word span, are either not correlated or only weakly correlated with reading ability (cf. Perfetti & Lesgold. 1977). The current paper proposes an alternative measure of working memory span that does correlate well with reading comprehension performance. The paper first discusses the nature of individual differences in working memory and presents the span test that was used to assess working memory capacity. Then it discusses how working memory capacity might influence two specific components of reading comprehension, retrieving facts and computing pronominal references. The first experiment shows that these two components correlate with the span measure. A second experiment shows that similar correlations are obtained when the span is assessed with a listening task.

While short-term memory traditionally has been conceived of as a passive storage buffer, the term working memory developed as a way to refer to a more active part of the human processing system (Newell, 1973). Working memory is assumed to have processing as well as storage functions; it serves as the site for executing processes and for storing the products of these processes (Baddeley & Hitch, 1974; LaBerge & Samuels, 1974). For instance in reading comprehension, the reader must store pragmatic, semantic and syntactic information from the preceding text and use it in disambiguating, parsing and integrating the subsequent text. Information can become part of working memory through several routes: it may be perceptually encoded from the text: it may be sufficiently activated so that it is retrieved from long-term memory: finally. it may be the output of a comprehension process (Carpenter & Daneman, Note 1). Information can also be lost from working memory, since its capacity is assumed to be limited (Miller, 1956; Simon, 1974). Information may be lost through decay or displacement. Decay occurs if the activation of information subsides to a subthreshold level with time (Collins & Loftus, 1975: Hitch, 1978; Reitman, 1974). Displacement occurs if additional structures are encoded, activated or constructed until the capacity is exceeded. An important aspect of information loss is the assumption that processes and structures compete for a shared limited capacity (Baddeley & Hitch, 1974: Case, 1978: Lesgold & Perfetti, 1979). Consequently, a task that has heavy processing requirements should decrease the amount of additional information that can be maintained. One way this could occur is if the execution of more demanding processes required more attention and hence consumed a larger proportion of the capacity otherwise available for storage. Another way is if the processes in the demanding task generated intermediate products that displaced the additional information.

The trade-off between processing and storage seems like a potential source of individual differences in reading comprehension (cf. also Perfetti & Lesgold, 1977). The better reader might have more efficient processes so that he/she effectively would have more capacity for storing and maintaining information. "More efficient processes" could have several interpretations. One minimal hypothesis is that the processes of good and poor readers differ only in some quantitative way. For example, a good reader may require fewer processes than a poor reader to perform exactly the same computation; in good readers, the intermediate steps might be eliminated in some or all of the stages such as decoding, lexical accessing, parsing, inferencing and integrating. Such efficiency would imply that the good reader would have fewer computational demands on working memory; hence, he would have more capacity for storing the necessary intermediate and final products of the reading process. More efficient processes would also be functionally faster and indeed, better readers are faster at reading-related tasks (Hunt, Frost & Lunneborg, 1973; Hunt, Lunneborg & Lewis, 1975; Jackson & McClelland, 1979; Perfetti & Lesgold, 1977). A speed advantage could interact with the decay of information from working memory since less of the preceding information would decay simply because of the passage of time. Hence, the more efficient processes of the good reader could be functionally equivalent to a larger storage capacity.

Contrary to this theory, the evidence so far suggests that working memory capacity may not differentiate good and poor readers. Studies using the standard digit span test

(Guyer & Friedman, 1975; Hunt, Frost & Lunneborg, 1973) or a probe digit span test (Perfetti & Goldman, 1976) have found no systematic differences between good and poor readers who were classified on the basis of a general reading comprehension test. Studies using letter-strings (Farnham-Diggory & Gregg, 1975; Rizzo, 1939) or similar sounding words (Valtin, 1973) as predictors of reading comprehension have been only slightly more successful. One explanation for the lack of correlation may be that digit span and word span tests do not sufficiently tax the processing component of working memory. The word span test, for instance, involves relatively simple processes such as rehearsal and access of common lexical items. A task with heavier processing demands might be needed to obtain a more marked trade-off between processing and storage. If the processing interfered with storage, the poor reader's less efficient processes would appear as equivalent to a smaller storage capacity.

One purpose of the present study was to devise a measure that taxed both the processing and storage functions of working memory. The processing and storage components of the test, which will be called the reading span test, involved the usual demands of sentence comprehension. An additional storage component required subjects to maintain and retrieve the final words of sentences. The format of the reading span test was somewhat similar to that of the traditional digit span and word span tests. The subject was given a set of sentences to read; at the end of the set, he attempted to recall the final word of each sentence. The number of sentences in a set was incremented from trial to trial and the subject's reading span was the maximum number of sentences he could read while maintaining perfect recall of the final words. If good readers use less processing capacity in comprehending the sentences, they should be able to produce more sentence final words than poor readers.

The span task was constructed so that its demands were compatible with the characteristics of working memory. First, the presentation times were short. The subject saw each sentence only for as long as it took to read it at a normal pace--approximately 5 seconds. As soon as one sentence was finished, the next was presented so that subjects weren't able to overtly rehearse the words. Second, the amount of information that the reader had to retain roughly matched the capacity attributed to working memory. For example, while reading the third sentence in a set, a reader would have to have sufficient processing and storage capacity for that sentence while retaining some representation that would allow him to retrieve the last words of the prior two sentences. This amount of information is consonant with reading models that assume working memory can accommodate several propositions from prior sentences while processing a subsequent sentence. Thus, the subject's reading span was taken as an index of his working memory capacity.

Relating working memory capacity to comprehension requires not only a measure of working memory span, but also an appropriate measure of comprehension. Traditional assessment techniques have relied on standardized reading comprehension tests, such as the Verbal Scholastic Aptitude Test (SAT). (Scores from SAT's were included in the present research also.) However, such global measures of comprehension are deficient from a theoretical point of view. Because such scores reflect a variety of subtasks, they are difficult to relate to any particular process. For that reason, the current research included two tests that tapped more specific components of comprehension: one required the reader to store and retrieve facts and the other required the reader to compute pronominal reference.

In the current experiment, the ability to encode, store and retrieve facts was assessed by asking the reader questions about simple facts in a short passage that had just been read. This kind of question-answering is a frequent component of reading comprehension tests (Carroll, 1972; Davis, 1944, 1968). One reason is that this task taps a skill that has both educational and practical importance. A second reason for such a test is that retrieving facts is a component of other comprehension processes. For example, to make an inference to relate some current information to a fact that was mentioned earlier in the text, the reader must retrieve the earlier information, as well as compute the relation. Thus, a test that requires simple fact retrieval after reading a short paragraph makes

demands on memory that may be comparable to the demands that are made during reading itself.

Both the initial encoding of facts and their subsequent retrieval involve working memory and could differentiate good and poor readers. Working memory capacity could influence both the duration that a fact remains in working memory and the probability that it is consolidated in long-term memory. In both cases, the better reader would have an advantage. A fact might persist longer in working memory for the better reader because his processing does not consume all of the available capacity. The fact will not be displaced as quickly. The good reader also might have an advantage in consolidating the fact in long-term memory. A larger processing capacity might allow more opportunities for integrating a particular fact into the general representation. The fact would be available during more of the subsequent processing so that later information could be related to it. Consequently, the integration process also would provide more retrieval routes for later accessing the fact. Finally, a reader with more efficient processes might have additional capacity to devote to rehearsal and consolidation. while the poorer reader would require all his processing capacity to perform the minimal computations. In summary, fact retrieval is one aspect of reading comprehension that could reflect differences in processing capacity.

A second component of comprehension that was measured required readers to compute pronominal references. This process may be related to fact retrieval, but it has special properties that seem closely linked to working memory capacity. When a writer uses a pronoun he is assuming that the referential concept is currently active in the reader's working memory or "foregrounded" (Chafe, 1972). Chafe compared the foregrounding of a concept to an actor who is introduced on stage during a play. To remain foregrounded, the concept must take some part in the action. If it doesn't, it has some probability of retreating to the wings, that is, of fading from the reader's working memory. Once the foregrounding of the concept is attenuated, the writer can no longer use the pronoun but must replace it with its referent noun. Chafe suggested that foregrounding might be attenuated after two sentences, although he admitted that this boundary is arbitrary and could be increased to an unspecified limit. The preceding analysis of working memory capacity suggests that the boundary might vary for individual readers. The distance over which a pronoun's referent could be computed might be partially dependent on the reader's working memory capacity. In particular, a larger capacity might mean a higher probability that the referent noun is still active. Hence, good readers might find such computations easier than poor readers. If the referent is no longer in working memory, the reader would have to institute a search of long-term memory. As in fact retrieval, the search might be more successful for better readers because they would have been more likely to have initially stored the original referent.

In summary, the central issue is whether the reading span test, as a measure of working memory capacity, correlates well with reading comprehension performance. Reading comprehension was assessed through a global score, the Verbal SAT, and two more specific comprehension tests. The first required readers to answer questions about facts given in narrative passages. The second required readers to answer questions that required pronominal reference: the task was one of identifying a pronoun's referent in passages that had increasing numbers of sentences intervening between the pronoun and its referent noun.

Experiment 1

Method

The subjects were given four tests: (1) a reading span test to measure the span of working memory, (2) a reading comprehension test that asked questions about facts and pronominal references, and (3) a traditional word span test.

Reading Span Test. Subjects had to read a series of sentences aloud at their own pace and recall the last word of each sentence. The test was constructed with 60 unrelated sentences, 13 to 16 words in length. Each sentence ended in a different word. Two examples are: When at last his eyes opened, there was no gleam of triumph, no shade of

anger. The taxi turned up Michigan Avenue where they had a clear view of the lake. Each sentence was typed on a single line across the center of an 8 x 5 inch index card. The cards were arranged in three sets each of 2, 3, 4, 5 and 6 sentences. Blank cards were inserted to mark the beginning and end of each set.

The experimenter showed one card at a time to the subject. The subject was required to read the sentence aloud. As soon as the sentence was read, a second card was placed on top of the first and the subject read the new sentence. The procedure was repeated until a blank card signalled that a trial had ended and that he was to recall the last word of each of the sentences in the order in which they had occurred. Subjects were given several practice items at the two sentence level before the test began. They were warned to expect the number of sentences per set to increase during the course of the test. The span test contained three sets each of 2, 3, 4, 5 and 6 sentences. Subjects were presented increasingly longer sets of sentences until they failed all three sets at a particular level. Testing was terminated at that point. The level at which a subject was correct on two out of three sets was taken as a measure of the subject's reading span. No subjects correctly recalled any set of cards at a higher level than their defined spans.

Reading Comprehension Tests. The subjects were given a series of passages to read and then at the end of each passage they were asked two questions; the first interrogated the referent of a pronoun mentioned in the last sentence and the second probed some other fact from the passage. Subjects were instructed to read each passage silently at a comfortable pace but to be prepared to answer questions about it. The passages were typed on separate sheets of paper and presented in random order. Subjects were given a sheet of cardboard to cover the lines of text as they completed reading them to prevent them from re-examining the text before answering the questions. After this task, they were questioned about their strategies; no subjects realized that they were being asked about pronominal reference after each paragraph.

There were twelve narrative passages of approximately 140 words in length. The fact question interrogated a simple fact with no restrictions as to the fact's location in the passage. The pronoun question always interrogated the referent of a pronoun that occurred in the last sentence. The passages were constructed so that the distance between the pronoun and its referent was systematically varied. The final sentence in each paragraph contained a pronoun, she, her, he, him or it. The antecedent noun occurred either 2, 3, 4, 5, 6 or 7 sentences prior to the final pronoun sentence. Varying the position of the antecedent noun prevented readers from adopting a strategy of attending to a particular sentence. Each of the six places was represented by two paragraphs. In each pair of paragraphs one referent noun was a common noun and one was a proper noun. In all cases, the referent of the pronoun was logically unambiguous. However, additional nouns of a similar class occurred in the sentences preceding the sentence containing the referent noun to make the task non-trivial. There was an attempt to make the referents relatively homogeneous in thematic importance since this factor may play a role in determining how long an item is foregrounded. Two sample passages and the fact and pronoun reference questions are presented in Table 1. In the Teenager Passage, the pronoun and its referent were separated by a distance of six sentences; in the Animal Passage, the distance was two sentences.

Insert Table 1 about here

Word Span Test. Subjects were required to recall sets of individual words. While the presentation was auditory, previous research has shown that auditory word span and visual word span correlate almost perfectly (Lyon, 1977). The test was constructed with 81 one-syllable common nouns that were as semantically and phonetically unrelated as possible. The procedure was similar to that used in the reading span test. The words were grouped in three sets each of 2, 3, 4, 5, 6 and 7 words. The experimenter presented the word sets orally to the subject at a rate of one word per second. Subjects were required to recall all of the words of a set in the exact order of presentation. They were warned to

expect the number of words per set to increase during the course of the test. Sets of increasing length were administered until a level was reached at which a subject failed all three sets. Phonetic confusions were very rare, but when a subject produced a word with one differing feature (such as gate for date) it was scored as correct, since it could have reflected an acoustic confusion. The level at which the subject was correct on two out of three sets was taken as a measure of the subject's word span.

The subject went through all four tests in the same order in a session lasting approximately 40 minutes. They were then asked their Verbal SAT score as a global measure of reading ability.

Subjects. The subjects were 20 Carnegie-Mellon University undergraduates who were enrolled in an introductory course in Psychology. They were all native speakers of English.

Results and Discussion

As shown in Table 2, the span test was correlated with the traditional assessment of comprehension, Verbal SAT scores, $r(18)=.59$, $p<.01$. The readers' SATs ranged from 400 to 710 with a mean of [illegible] (S.D.=79.8). The span test was even more closely related to performance on the two specific tests of comprehension, the fact questions and the pronominal reference questions; the correlations were $r(18)=.72$ and .90, respectively, and $p<.01$ for both. The reading span for the 20 readers varied from 2 to 5 with a mean of 3.15 (S.D.=.93). Readers with smaller spans performed much worse than readers with larger spans on both tests. For example, the five readers with span size 2 correctly answered only 8.2 out of the 12 fact questions and 5.4 out of the 12 pronoun reference questions. By contrast, the six readers with span sizes 4 and 5 answered 11 out of the 12 fact questions and 9.7 of the 12 pronoun reference questions. The mean performance overall was 9.4 for the fact questions and 7.4 for the pronoun reference questions.

The traditional word span measure in this study, as in many previous studies, was not significantly correlated with any comprehension measure, whether Verbal SAT, $r(18)=.35$, fact questions, $r(18)=.37$, or pronominal reference, $r(18)=.33$, as shown in Table 2. Moreover, the correlations between word span and the two specific comprehension tests were significantly lower than the correlations between reading span and these tests, $t(19)=2.99$, $p<.01$, for the fact questions and $t(19)=5.92$, $p<.01$, for the pronominal reference questions. Scores on the traditional word span test for the 20 subjects varied from 4 to 6; the mean was 5.15 (S.D.=.67), higher than the measure of working memory span. The two span tests were moderately correlated themselves, $r(18)=.55$, $p<.01$.

Insert Table 2 about here

The results for the pronoun reference questions showed that there was a close correspondence between the reader's span and the distance over which he/she could correctly answer the question. The results for readers with different spans for small, medium and large pronoun-referent noun distances is shown in Figure 1. There was no significant differences in computing a common versus proper referent noun ($t(19)<1$) so the two types of nouns were not treated separately in the analyses. An analysis of variance with four spans and six distances showed that readers with larger reading spans performed better, $F(3,16)=27.56$, $p<.01$, and readers tended to do better on passages with smaller distances, $F(5,80)=37.10$, $p<.01$. In addition, span interacted with the distance over which a reader could make a pronominal inference, $F(15,80)=2.10$, $p<.05$. As shown in Figure 1, readers with a span of 2 or 3 had difficulty with noun-pronoun distances greater than 2 or 3. Readers with a span of 4 made numerous errors only when the distance was greater than 5 sentences. Span 5 subjects had no errors even for distances of 6 and 7. The analysis of variance used a least squares formula to handle the unequal n's in the various spans. Post hoc individual comparisons indicated that readers with different spans were significantly different from each other except that the span 2 and span 3 readers' performances did not differ significantly.

In summary, several aspects of the results support the hypothesis that the reading span task is related to working memory capacity. In particular, the limits of performance in the span task are consistent with a limited working memory. It is interesting to note that a span of 5 sentence final words is the best performance observed in this experiment or in the experiments that have been run subsequently (with approximately 100 subjects in all). Moreover, subjects' retrospective reports about how they performed the task suggests that they actively tried to retain the sentence-final words in working memory. Subjects sometimes reported trying to rehearse the words mentally while reading the next sentence aloud. Also, subjects sometimes reported forming associations among sentence-final words. Finally, it was apparent during some of the recall trials in the span test that subjects sometimes used reconstructive processes. When the subject was unable to produce the sentence-final word, he/she might attempt to reconstruct the sentence on the basis of whatever 'gist' had been retained. In these reconstructions, subjects sometimes made systematic errors by choosing a word that had been present in the sentence but wasn't the final word. These informal observations and the subjects' retrospections suggest that the reading span task was successful in taxing the processing and storage capacity of working memory. Of primary importance, of course, is the fact that the span task correlates with three measures of reading comprehension. The theory proposes that working memory capacity is the source of the correlation and hence, an important source of individual differences in reading.

Experiment 2

While the role of working memory in comprehension has been described in the context of reading, the same processes could occur in listening comprehension as well; no specifically visual processes are central to the present argument. Experiment 2 examined working memory capacity for spoken as well as written verbal material to determine the relation between reading span and listening span and their correlations with reading and listening comprehension.

To assess listening span, the test was modified so that it was suitable for both reading and listening. The modified span test consisted of sentences to be individually verified as true or false. Then, at a signal, the subject recalled the last word in each sentence. This span measure was used with silent reading, oral reading and listening. These span measures were then compared to silent reading comprehension and listening comprehension measures. The verification component was included in the span measure to assure that subjects processed the entire sentence and did not concentrate only on the sentence-final word.

The oral component of Experiment 2 also ruled out a reading speed interpretation of Experiment 1, that is, that the good readers are superior to bad readers solely because they are faster and therefore had shorter retention intervals between the reading of a critical word and recall. In the present experiment, the listening span test and listening comprehension test were presented at a constant rate and still revealed systematic individual differences.

This experiment also examined evidence for qualitative differences in the errors made by readers with different spans. The major analysis classified readers' erroneous responses to fact questions and pronominal reference questions according to how consistent they were with the gist of the passage. The issue was whether readers with large and small spans make different kinds of errors.

Method

The subjects were given three working memory span tests: (1) an oral reading span test, (2) a silent reading span test, and (3) a listening span test. These were followed by two comprehension tests, one involved silent reading and the other one, listening.

<u>The Reading Span and Listening Span Tests</u>. Subjects had to read or listen to a sentence and answer true or false. Sentences were presented in a set and at the end of a set, the reader had to recall the last word of each sentence. The true-false component was

included to ensure that subjects processed the entire sentence and did not just concentrate on the final words, a possible strategy if subjects had to only listen or silently read a sentence. While this is not a possible strategy when reading orally, the true-false component was also included in the oral reading span test to keep all three span tests comparable.

The span tests were constructed from a set of 220 different sentences. Each was 9 to 16 words in length, and ended in a different word. The sentences were taken from general knowledge quiz books and covered a wide number of knowledge domains such as the biological and physical sciences, literature, geography, history and current affairs. Half were true and half were false. Care was taken to select statements that were of moderate difficulty, for example, (1) You can trace the languages English and German back to the same roots. (2) The Supreme Court of the United States has eleven justices.

The subjects read or listened to the sentence and then had 1.5 sec in which to answer true or false before the next sentence was presented. In the reading versions, the end of a trial was signalled by a blank card, in the listening version, by a tone. Subjects were given several practice items at the two sentence level before the test began. If a subject had difficulty verifying the sentence in 1.5 sec, he was urged to try to answer correctly and if he didn't know the answer, to give the most plausible answer. All the subjects mastered the procedure by the end of the practice sesion. Whether or not subjects verified the sentences correctly was ignored. At the end of the experiment subjects were questioned about their strategies; they all believed the true-false aspect was relevant.

The silent and oral reading span tests contained three sets each of 2, 3, 4, 5 and 6 sentences. As in Experiment 1, subjects were presented increasingly longer sets of sentences until they failed to recall the sentence-final words of all three sets at a particular level. Testing was terminated at this point. The level at which a subject was correct on two out of three sets was taken as a measure of the subject's reading span. If the subject was correct on only one set at a particular level he was given a credit of .5. For example, if a subject was correct on two out of the three 4-sentence sets he was assigned a span of 4. If he was correct on only one of the three sets he was assigned a span of 3.5. Because the test proved to be so difficult, the subject was given credit for any set for which he recalled all sentence-final words, irrespective of the order of recall. In the listening span test there were five sets each of 2, 3, 4, 5 and 6 sentences. Subjects were defined as being at the span level at which they were correct on three out of the five sets. If they were correct on two out of the five sets they were given a credit of .5.

The Comprehension Tests. The materials and procedure were similar to those in Experiment 1. The only change was that subjects were asked four rather than two questions at the completion of each passage. The first question required the identification of the pronominal referent. The second and third questions interrogated other facts from the passage. The fourth question asked for a title for the passage.

There were 24 passages which were randomly divided into two groups of 12 each, with the constraint that 2 out of the 4 passages at each pronoun-noun distance level was assigned to a group. Half the subjects read the first group of passages and listened to the second group; the other subjects read the second group and listened to the first. The listening comprehension paragraphs and questions were tape recorded by a male speaker at the speed of normal speech.

All subjects were given the three span tests followed by the two comprehension tests in a session lasting approximately one hour. To control for order effects, half the subjects were given the reading versions of the span and comprehension tests before the listening ones, while the remaining subjects received the tests in the opposite order. Finally, the subjects' Verbal SAT scores were requested as an additional measure of reading ability.

Subjects. The subjects were 21 Carnegie-Mellon University undergraduates who were enrolled in an introductory Psychology course. They were all native speakers of English.

Results and Discussion

The results of Experiment 2 corroborate the findings of Experiment 1 by demonstrating high correlations between the measures of working memory span and the measures of reading comprehension. As shown in Table 3, readers with large oral reading spans were better at fact questions and pronoun reference questions, $r(19)=.81$ and $.84$, $p<.01$, respectively. These correlations are similar to the .72 and .90 found in Experiment 1 with an oral reading span task that did not require verification. Similarly, readers with large silent reading spans were better at fact questions and pronoun reference questions $r(19)=.74$ and $.86$, $p<.01$, respectively. Moreover, both spans show significant correlations with Verbal SAT scores, $r(19)=.55$, $p<.01$, $r(19)=.49$, $p<.05$, compared to $r(18)=.59$ in Experiment 1. The subjects' SATs ranged from 410 to 760 with a mean of 600 (S.D.=75).

Listening span measures are almost as good at predicting reading comprehension as reading span measures. Subjects with larger listening spans were better at answering questions about facts and pronominal references that had been read, $r(19)=.67$ and $.72$, $p<.01$, respectively. While the correlations are somewhat lower than those obtained with reading spans, they are still impressive. Finally, the listening span measure correlated as well with the Verbal SATS as did the other spans, $r(19)=.53$, $p<.01$.

Insert Table 3 about here

Not surprisingly, the three span measures were highly correlated; the oral reading span correlated .88 with the silent reading span and .80 with the listening span, and the silent reading span correlated .75 with the listening span, all $p<.01$. The absolute range of the spans was also similar. The oral reading span scores of the 21 subjects varied from 1.5 to 5, with a mean of 2.76 (S.D.=.80). The silent span varied from 1.5 to 4, with a mean of 2.38 (S.D.=70). Finally, the listening span scores varied from 2 to 4.5, with a mean of 2.95 (S.D.=.72).

All three span measures did reasonably well in accounting for listening comprehension, although the listening span measure was slightly better. As shown in Table 3, the correlations for listening fact questions ranged from .42 to .47, $p<.05$. These correlations are lower than those obtained for reading fact questions. This appeared to be due to three subjects who performed better than would be predicted by their span task, rather than due to some question that was systematically easier in listening than reading. The correlations obtained for listening pronoun inference questions are much more like those obtained in reading, ranging from .71 to .85, $p<.01$.

The relatively high correlations between reading and listening performance suggests that a significant amount of individual difference is common to both reading and listening. Particularly for adult readers, such as the present university population, visual encoding processes may not be the bottleneck in reading. Rather, comprehension processes that tax working memory may be responsible for difficulties that subjects have in either reading or listening. This result is somewhat similar to that of Sticht (1972) for a group of Army personnel. He found that subjects who had problems reading more difficult material also had problems when it was presented orally. Both sets of results suggest that an individual differences theory that stresses visual processes, such as the time to access letter codes from visual material, may not capture a significant source of variance that is common to both reading and listening comprehension.

Performance on the pronominal reference questions is shown in Figures 2 and 3 for the reading and listening conditions. The same analyses of variance as in Experiment 1 were applied to the data. The overall results and the patterns of performance were similar for both the reading and listening conditions. As in Experiment 1, subjects with large silent reading and listening spans performed better, $F(4,16)=12.14$, $p<.01$, and $F(5,15)=8.33$, $p<.01$, respectively. Subjects tended to do better on the written and spoken passages with shorter distance, $F(2,32)=11.25$, $p<.01$ and $F(2,32)=9.16$, $p<.01$, respectively. Unlike Experiment 1, the interactions between span and distance for both the reading and listening data were not significant. In both Experiments 1 and 2, the subjects with the largest working memory

spans performed at asymptote. The main difference was between the slopes for the remaining spans in Experiment 1 (Figure 1) and those in Experiment 2 (Figures 2 and 3). In Experiment 1, even the small span subjects made relatively few errors on the short distance passages. However, in Experiment 2, all subjects with silent reading or listening spans less than 3.5 made errors at the shortest distances, with the percentage of errors increasing inversely with span size. In other words, in Experiment 2, subjects with large spans had an advantage over low span subjects at retrieving referents at all distances, not just when the distance was large. Furthermore, in Experiment 1 the performance of subjects with different spans began dropping at different rates; in Experiment 2 performance dropped at a parallel and more gradual rate.

Insert Figures 2 and 3 about here

In order to examine possible qualitative differences between good and poor readers an analysis was made of their errors in the comprehension tests. The analysis was done on the four conditions: the reading fact questions, the reading pronoun reference questions, the listening fact questions and the listening pronoun reference questions. For the reading analysis, subjects were classified into large and small silent reading spans. The cutoff point was designated as the span interval at which the largest decrement in performance occurred. This occurred between spans 2.5 and 3. Readers with larger spans made an average of 5.4 errors out of 24 fact questions and 2.5 out of the 12 pronoun reference questions. Subjects with smaller spans made an average of 11.3 errors out of the 24 fact questions and 6.2 of the 12 reference questions. For the listening task, the cutoff occurred between 3 and 3.5. Subjects with large listening spans made an average of 6.6 errors out of the 24 fact questions and 1.4 errors out of the 12 pronoun reference questions. Subjects with small listening spans made an average of 8.8 errors out of 24 fact questions and 5.1 out of 12 pronoun reference questions. The erroneous responses to fact and pronoun reference questions in the reading and listening tasks were classified into several categories: reasonable substitutions, incomplete answers, foils from the passage, fabrications, confusions and don't remembers.

The first category of errors were reasonable errors; these included reasonable substitutions and incomplete but partially correct answers. A reasonable substitution was a superset for the correct answer. An example from the Teenager passage in Table 1 for the question, <u>Where was the gang sitting?</u>, would be the answer <u>the restaurant</u> as a substitution for <u>the Grill</u>. An incomplete answer was one in which subjects gave only part of an answer. An example from the Animal passage in Table 1 for the question, <u>Where was the emergency meeting held?</u>, would be the answer the <u>river</u> rather than <u>the river clearing</u>. The incomplete answers category was irrelevant for the pronoun reference questions since the response consisted of a single referent noun. Reasonable substitutions and incomplete answers were considered together for fact questions. The two types were collapsed because they are both very consistent with the gist of a passage and because they were distributed similarly within the small and large span groups.

Another category included foil errors and fabrications. A foil was an item that occured in the passage and had features similar to the correct answer. An example in the Teenager passage for the question <u>Who put the money in the juke box?</u>, would be the answer <u>Archie</u> instead of <u>Robbie</u>. Fabrications were items that did not occur in the passage although they resembled the correct answer in features such as number and gender. An example for the question about who put the money in the juke box would be the response <u>Tom</u> because <u>Tom</u> did not occur in the passage. Foils and fabrications were collapsed because they had similar distributions within small and large span groups.

The last two error types were "don't remembers" and confusions. Confusions indicated a fundamental misunderstanding. An example from the Teenager passage for the question, <u>Who was 'all shook up' and serious over the music?</u>, would be the answer <u>the narrator</u>; such an error reflects that the subject failed to process the major distinction made by the passage about the narrator and his best friend.

The error analysis showed that subjects with larger spans made less serious errors. The distributions for subjects with large and small silent reading spans and large and small listening spans is shown in Table 4. For both the reading and listening fact questions, subjects with larger spans made different sorts of errors than subjects with smaller spans, $\underline{X}$ (3)=13.20, $\underline{p}$<.01, and $\underline{X}$ (3)=13.70, $\underline{p}$<.01, respectively. The subjects with larger spans made a larger percentage of the reasonable errors (reasonable substitutions and incompletes); subjects with smaller spans made a larger percentage of the types of errors that indicated more serious problems (the confusions and don't remembers). The same trend was present for the pronoun reference questions in the reading and listening conditions $\underline{X}$ (3)=10.30, $\underline{p}$ <.01, and $\underline{X}$ (3)=37.2, $\underline{p}$<.01, respectively.

Insert Table 4 about here

Not only do good readers make fewer errors than poor readers but their errors are more likely to reflect at least some understanding of the passages. On the other hand, when poor readers or listeners fail to retrieve a fact or a pronoun's referent, their errors frequently reflect a more fundamental misunderstanding of the passage. Their errors might contradict the whole theme of the passage. This was particularly evident in the confusion errors for pronoun reference questions. A typical example in the Teenager passage was the answer to the question Who was "all shook up" and serious over the music? the correct answer was Wayne (the narrator's best friend); but some subjects answered the narrator. While none of the subjects with large spans made this error, six subjects with small spans did. Such an error suggests that the reader or listener had misunderstood the central theme, which was the narrator's scorn for the seriousness with which his friends reacted to the music.

Another indication of a qualitative difference among subjects appeared in the themes that were given for the passages. The titles that subjects provided for each passage in the reading and listening comprehension tests were scored from 0-2 according to how well they expressed the main idea. A score of 0 was assigned to a title that represented an unimportant detail of the passage or that denoted an incorrect interpretation of the passage; examples of titles from the teenager passage in Table 1 were The Boys' Night Out and The Malt Shop. A score of 1 was assigned to a title that captured the theme of a passage at a very general level or that represented an incomplete theme; an example from the same passage was the title Teenage Boys. A score of 2 was given to a title that was a more detailed account of the passage or that succintly expressed the author's intentions or moral; examples were My Objections to Rock Music and Taking Things in Their Proper Perspective.

Subjects with larger spans were better at abstracting a theme from a written or spoken narrative passage. In the reading comprehension task, the mean rating for subjects with large silent reading spans was 16 out of 24; for subjects with small spans the mean was 11.8 out of 24. In the listening task, the mean rating for subjects with large listening spans was 16.7 out of 24; for subjects with small spans the mean was 12.1 out of 24. The ability to abstract a theme from written narratives correlated significantly with silent reading span $\underline{r}$(19)=.71, $\underline{p}$<.01. Similarly, the ability to abstract a theme from spoken narratives correlated significantly with listening span, $\underline{r}$(19)=.82, $\underline{p}$<.01. Again, abstracting themes from written and from spoken verbal material seemed to share common skills: silent reading span correlated with listening theme abstraction, $\underline{r}$(19)=.57, $\underline{p}$<.05; listening span correlated with reading theme abstraction $\underline{r}$(19)=.69, $\underline{p}$<.01.

General Discussion

The high correlations between reading span and the various comprehension measures are striking; the reading span task succeeds where previous short-term memory measures have failed. The argument has been that the span task reflects working memory capacity and that this capacity is a crucial source of individual differences in language comprehension. While more promising than previous individual differences analyses, the present studies are still correlational and more evidence will be needed to show precisely

how and when working memory capacity limits the comprehension process.

One useful experimental approach might expand on the design of the pronominal reference task. To successfully compute a pronominal referent, the reader must retrieve prior information, in this case, the referent, and link it to the pronoun. All integrative computations involve relating some current information to previous information. The previous information might have been explicitly stated or it might have to be inferred. It seems reasonable that the computation will be easier if the required prior information is active in working memory and harder if it must be searched for and retrieved from long-term memory. The present theory predicts that readers with smaller spans are less likely to have the prior information active in working memory. In the current study, this lowered probability manifested itself in higher error rates. However, it might also manifest itself in longer response times, assuming that the poorer reader attempts to search long-term memory, and in overt patterns of eye fixations including pauses and regressions. These are issues that we are currently investigating.

The present paper has stressed quantitative differences in working memory capacity as the source of individual differences. The most parsimonious assumption seems to be that readers are making the same computations, but that they differ in the speed or efficiency with which they can make those computations. However, the analysis of readers' errors also indicated some qualitative differences among readers: high and low span readers do not make the same kinds of errors. One theoretical explanation might be that the processes of good and poor readers are qualitatively different. Poor readers may be doing fundamentally different things than good readers and the different processes also may be less efficient. An alternative explanation, and one to be pursued here, is that qualitative differences can emerge from quantitative differences.

One example of how differences in capacity could result in qualitative differences in processing is in the chunking process. This process recodes concepts and relations into higher order units. One prerequisite for chunking is that each of the individual concepts be present simultaneously in working memory (Shiffrin & Schneider, 1977). Consequently, the more concepts there are to be organized into a single chunk, the more working memory will be implicated. Although the process of forming rich chunks would temporarily strain working memory capacity, these chunks would have a quantitative payoff. The recoding of many concepts and relations into a single chunk would have the economizing effect of reducing the load on working memory and hence would increase the functional working capacity for subsequent processing.

The chunks of a good reader might differ qualitatively from a poor reader. The good reader has more functional working memory capacity available for the demands of chunking. He is more likely to have more concepts and relations from preceding parts of the text still active in working memory. Consequently, he should be able to detect more interrelations among these concepts, and to note their relative importance. The good reader's chunks should be richer, and more coherent, and contain different information. The presence of different interrelations could subsequently allow different inferences and generalizations to be drawn.

Good readers might capitalize on their chunking efficiency in the comprehension and working memory span tasks. In the comprehension test passages, the chunks of the good reader would correspond to the interrelations of clauses and sentences that form superordinate discourse units. Such chunks might aid readers in the storage and retrieval of facts and themes. In the span test, which consisted of unrelated sentences, the role of chunking strategies is less obvious. The chunks might be the interrelations among the sentences of a set or even the interrelations of propositions within a single test sentence. As mentioned before, there is some suggestion from the verbal reports of our subjects that both types of chunking occurred.

The present theory that working memory capacity is the crucial source of individual differences in reading was based on a circumscribed population, that is, on a group of bright university students. The mean Verbal SAT score of the group was above the national mean and presumably even the poorest readers were functioning adequately with

undergraduate level reading materials. The theory does not address itself to other possible sources of individual differences in reading, such as perceptual disabilities and motivational factors, that may be characteristic of the extremely poor or immature readers.

This research on individual differences has a parallel to the research on memory development. The argument has been that a major difference between good and poor readers is the efficiency of their processing, rather than static memory capacity. Similarly, developmental research suggests that children and adults differ in processing capacity and that there is little difference in their static capacities. If adults are prevented from using special strategies and if the material is equated for familiarity, children and adult memory spans are very similar (Case, 1978; Chi, 1976; Huttenlocher & Burke, 1976). But children have much slower and less efficient processes. The more of these processes the task requires, the more marked are the developmental differences. Thus both individual and developmental differences are most evident when the subjects differ in the efficiency with which they can execute required processes.

The converse of this argument is that if subjects are equally adept at a particular process, there should be few individual differences in a span task that requires that process. One example of such a process is simple counting. In a counting span task that paralleled the reading span task, Case, Kurland & Daneman (1979) had adults and children count arrays and recall the number of items in each array in a set. The number of arrays increased until the subject could no longer recall accurately. Adults seemed to be equally good at counting. There were only small differences in the speed of counting, a measure of processing efficiency, and in the counting span. Moreover, there was no correlation between the speed and span measures. Individual differences among adults were found only when their processing efficiency was impaired by having them count in an artificial language. By contrast, children are not equally efficient in the process of simple counting. For them, there was a high correlation between their counting span and counting speed.

The asymptotic efficiency of a simple skill like counting highlights the lack of asymptotic performance in reading. If one assumes that a reading span of 5 is asymptotic, the conclusion is that most readers (even a group of bright undergraduate college students) are not performing at ceiling. Moreover, the results of Experiment 2 show that this conclusion is not peculiar to reading. There were significant individual differences in listening comprehension as well. The interesting implication is that both reading and listening comprehension are not easy for adults and that there may be potential room for improvement well into adulthood.

Reference Note

Carpenter, P. A., & Daneman, M. Lexical retrieval and error recovery in reading: A model based on eye fixations. Mimeograph manuscript, Carnegie-Mellon University, Pittsburgh, PA, 1979.

References

Baddeley, A. D., & Hitch, G. Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 8). New York: Academic Press, 1974.

Carroll, J. B. Defining language comprehension: Some speculations. In R. O. Freedle & J. B. Carrol (Eds.), Language comprehension and the acquisition of knowledge. Washington D. C.: V. H. Winston & Sons, 1972.

Case, R. Intellectual development from birth to adulthood: A Neo-Piagetian interpretation. In R. Siegler (Ed.), Children's thinking: What develops?. Hillsdale, NJ: Lawrence Erlbaum Associates, 1978.

Case, R., Kurland, M., & Daneman, M. Operational efficiency and the growth of M-space. Paper presented at the Society for Research in Child Development, San Francisco, March, 1979.

Chafe, W. L. Discourse structure and human knowledge. In R. O. Freedle & J. B. Carroll (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: V. H. Winston & Sons, 1972.

Chi, M. T. H. Short-term memory limitations in children: Capacity or processing deficits? Memory and Cognition, 1976, 4, 559-580.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Davis, F. B. Fundamental factors in reading. Psychometrika, 1944, 9, 185-197.

Davis, F. B. Research in comprehension in reading. Reading Research Quarterly, 1968, 3, 499-545.

Farnham-Diggory, S., & Gregg, L. W. Short-term memory function in young readers. Journal of Experimental Child Psychology, 1975, 19, 279-298.

Guyer, B. L., & Friedman, M. P. Hemispheric processing and cognitive styles in learning-disabled and normal children. Child Development, 1975, 46, 658-668.

Hitch, G. J. The role of short-term working memory in mental arithmetic. Cognitive Psychology., 1978, 10, 302-323.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 7). New York: Academic Press, 1973.

Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? Cognitive Psychology, 1975, 7, 194-227.

Huttenlocher, J., & Burke, D. Why does memory span increase with age? Journal of Verbal Learning and Verbal Behavior, 1976, 8, 1-31.

Jackson, M. D., & McClelland, J. L. Processing determinants of reading speed. Journal of Experimental Psychology: General, 1979, 108, 151-181.

Just, M. A., and Carpenter, P. A. A theory of reading comprehension. To appear in C. Perfetti and A. Lesgold (Eds.) Interactive processes in reading. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Lesgold, A. M., & Perfetti, C. A. Interactive processes in reading comprehension. Discourse Processes, 1978, 1, 323-336.

Lyon, D. R. Individual differences in immediate serial recall: A matter of mnemonics? Cognitive Psychology, 1977, 9 403-411.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Newell, A. Production systems: Models of control structures. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

Perfetti, C. A., & Goldman, S. R. Discourse memory and reading comprehension skill. Journal of Verbal Learning and Verbal Behavior, 1976, 14, 33-42.

Perfetti, C. A., & Lesgold, A. M. Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Reitman, J. S. Without surreptitious rehearsal information in short-term memory decays. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 365-377.

Rizzo, N. D. Studies in visual and auditory memory span with specific reference to reading disability. Journal of Experimental Education, 1939, 8, 208-244.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II Perceptual learning, automatic attending and a general theory. Psychological Review, 1977, 84, 127-190.

Simon, H. A. How big is a chunk? Science, 1974, 183, 482-488.

Sticht, T. G. Learning by listening. In R. O. Freedle & J. B. Carroll (Eds.), Language comprehension and the acquisition of knowledge. Washington D. C.: V. H. Winston & Sons, 1972.

Valtin, R. Reports of research on dyslexia in children. Paper presented at International Reading Association, Denver, 1973. (ERIC Document Reproduction Service No. ED 079713)

Footnote

We thank Marcel Just for his comments on the paper. The research was partially supported by a Canada Council Fellowship to Meredyth Daneman and by grant NIE 77-0007 from the National Institute of Education to Patricia Carpenter.

Requests for reprints should be sent to Patricia A. Carpenter, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pa. 15213.

Table 1
Sample Passages and Questions from the Comprehension Tests

Teenager Passage (Distance 6)

Sitting with Richie, Archie, Walter and the rest of my gang in the Grill yesterday, I began to feel uneasy. Robbie had put a dime in the juke box. It was blaring one of the latest "Rock and Roll" favorites. I was studying, in horror, the reactions of my friends to the music. I was especially perturbed by the expression on my best friend's face. Wayne looked intense and was pounding the table furiously to the beat. Now, I like most of the things other teenage boys like. I like girls with soft blonde hair, girls with dark curly hair, in fact all girls. I like milkshakes, football games and beach parties. I like denim jeans, fancy T-shirts and sneakers. It is not that I dislike rock music but I think it is supposed to be fun and not taken too seriously. And here he was, "all shook up" and serious over the crazy music.

Questions:

Pronoun	(1) Who was "all shook up" and serious over the music?
Fact	(2) Where was the gang sitting?
Fact	(3) *Who put money in the juke box?
Theme	(4) *Provide a title for the passage that captures its theme.

Table 1 Continued
Animal Passage (Distance 2)

It was midnight and the jungle was very still. Suddenly the cry of a wolf pierced the air. This anguished note was followed by a flurry of activity. All the beasts of the jungle recognized that an urgent meeting had been summoned by the lion, their king. Representatives from each species made rapid preparations to get to the river clearing. This was where all such emergency assemblies were held. The elephant and tiger were the first to arrive. Next came the gorilla, panther and snake. They were followed by the owl and the crocodile. The proceedings were delayed because the leopard had not shown up yet. There was much speculation as to the reasons for the midnight alarm. Finally he arrived and the meeting could commence.

Questions:

Pronoun	(1) Who finally arrived?
Fact	(2) Where were these emergency meetings held?
Fact	(3)* What broke the stillness of the night?
Theme	(4)* Provide a title for the passage that captures its theme.

*These questions were included only in Experiment 2.

Figure Captions

Figure 1. The percentage of correct responses to the pronoun reference questions as a function of the distance between the pronoun and the referent noun. The curves represent the various reading spans in Experiment 1.

Figure 2. The percentage of correct responses to the pronoun reference questions in the reading comprehension task as a function of the distance between the pronoun and its referent noun. The curves represent the various silent reading spans in Experiment 2.

Figure 3. The percentage of correct responses to the pronoun reference questions in the listening comprehension task as a function of the distance between the pronoun and its referent noun. The curves represent the various listening spans in Experiment 2.